Proposition d'un cadre pour l'analyse automatique, l'interprétation et la recherche interactive d'images de bande dessinée

Clément Guérin

▶ **To cite this version:**

Clément Guérin. Proposition d'un cadre pour l'analyse automatique, l'interprétation et la recherche interactive d'images de bande dessinée. Traitement des images [eess.IV]. Université de La Rochelle, 2014. Français. NNT : 2014LAROS024 . tel-01175501

HAL Id: tel-01175501

https://theses.hal.science/tel-01175501

Submitted on 10 Jul 2015

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

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

UNIVERSITÉ DE LA ROCHELLE

ÉCOLE DOCTORALE S2IM

LABORATOIRE : L3I

THÈSE présentée par :

Clément GUÉRIN

soutenue le : **24 novembre 2014**
pour obtenir le grade de : **Docteur de l'université de La Rochelle**
Discipline : **Informatique et Applications**

Proposition d'un cadre pour l'analyse automatique, l'interprétation et la recherche interactive d'images de bande dessinée.

RAPPORTEURS	**Rokia MISSAOUI**	Professeur, LARIM, Université du Québec en Outaouais
	Céline HUDELOT	Maître de conférences HDR, MAS, École Centrale de Paris
EXAMINATEURS	**Amedeo NAPOLI**	Directeur de recherche CNRS, LORIA, Université de Lorraine
	Jean-Philippe DOMENGER	Professeur, LABRI, Université Bordeaux 1
	Josep LLADÓS	Professor, CVC, Universitat Autònoma de Barcelona
INVITÉS	**Sébastien FERRÉ**	Maître de conférences HDR, IRISA, Université de Rennes 1
	Jean-Christophe BURIE	Professeur, L3i, Université de La Rochelle
DIRECTION	**Karell BERTET**	Maître de conférences HDR, L3i, Université de La Rochelle
	Arnaud REVEL	Professeur, L3i, Université de La Rochelle

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

REGION
Poitou
Charentes
la démocratie participative

la
Charente
Maritime

Informatique
Image
Interaction
Univ. La Rochelle

Remerciements

Les éléments présentés dans ce manuscrit sont l'aboutissement de réflexions et de travaux débutés en 2011 et menés durant trois ans. De mes premiers pas dans le monde de la recherche, effectués en fin de master, à la soutenance de ma thèse de doctorat, j'ai eu la chance et la joie de croiser le chemin d'un nombre incroyable de gens formidables. Ils m'ont tous aidé, à leur manière, à un moment ou à un autre, à arriver au bout de ces travaux et je me dois maintenant de les remercier, en croisant les doigts pour n'oublier personne.

Mes premiers remerciements vont, tout naturellement, à mes encadrants, Karell et Arnaud. D'une présence et d'une disponibilité sans faille tout du long de cette aventure, vous avez toujours su m'encourager et me remotiver lorsqu'il y en avait besoin. J'ai, grâce à vous, pu travailler et m'intéresser à un grand nombre de choses très différentes, toujours dans une excellente dynamique et, pour tout cela, je vous remercie sincèrement. Je tiens également à présenter mes remerciements à l'ensemble des membres de mon jury pour avoir accepté de faire le déplacement afin d'examiner ces travaux. Merci à, dans l'ordre alphabétique, Jean-Christophe Burie, Jean-Philippe Domenger, Sébastien Ferré, Josep Lladós et Amedeo Napoli. J'adresse également des remerciements tout particuliers à Céline Hudelot et Rokia Missaoui, ayant généreusement accepté de relire et d'évaluer ce document. Merci à vous pour votre travail de relecture et pour vos remarques avisées m'ayant permis d'améliorer – et d'ajouter un nombre certain d'accents circonflexes à – cette thèse.

J'ai effectué ma thèse au sein du laboratoire L3i dont je remercie l'équipe directrice pour la confiance qu'elle m'a accordée quant à la réalisation de ces travaux. Merci également à Chris, Antoine et Norbert, mes collègues du projet « eBDtheque », dans lequel j'ai pu évoluer ces dernières années. J'en profite pour remercier également tous mes collègues et amis et, d'une manière plus générale, toutes les personnes que j'ai eu la chance de côtoyer ces dernières années, pour les réflexions stimulantes qui permettent d'avancer et les moments de détentes qui permettent de souffler. Doctorants, ingénieurs, administratifs, professeurs, stagiaires du laboratoire et d'ailleurs, merci à vous.

Enfin, *last but not least*, merci à toute ma famille, à mes parents, mes grands-parents, vous qui m'avez donné l'opportunité de faire des études et de les poursuivre jusqu'à ce niveau. Sans vous je n'en serai pas là. Merci. Et, bien entendu, merci à toi Pauline de m'avoir toujours compris, soutenu et de m'avoir supporté durant ces derniers mois, pas évidents, intégralement passés à lire et à écrire. La rédaction de cette sacrée thèse s'achève avec cette phrase, c'est fini.

Table des matières

1 Introduction générale **1**
1.1 La bande dessinée,... 1
1.2 ...un patrimoine mondial... 2
1.2.1 La bande dessinée franco-belge 2
1.2.2 Le comics nord-américain 3
1.2.3 Le manga japonais 4
1.2.4 Le webcomic 5
1.3 ...en pleine mutation... 5
1.3.1 L'exemple des autres médias 5
1.3.2 Le cas de la bande dessinée 6
1.4 ...générant de nouveaux défis. 8

I Des connaissances et des outils pour une analyse visuelle et sémantique de la bande dessinée. **11**

2 Représentation des connaissances **13**
2.1 Introduction 13
2.2 L'exemple du web sémantique 14
2.2.1 Du web des données au web social 14
2.2.2 Le web sémantique 15
2.3 Les ontologies 19
2.3.1 Définition 19
2.3.2 Composition 20
2.4 Les Logiques de Description (LD) 23
2.4.1 Introduction 23
2.4.2 Différentes logiques fondatrices 23
2.4.3 Une famille de logiques 28
2.4.4 Sémantique 31
2.4.5 Raisonnement 31
2.5 Conclusion 33

3 Analyse de la bande dessinée **35**
3.1 Introduction 35
3.2 Des codes graphiques établis 36
3.2.1 Un univers décrit en cases 36
3.2.2 Restitution de la bande son 41
3.3 Point de vue traitement d'images 46
3.3.1 Extraction des cases 46
3.3.2 Extraction des bulles 47

3.3.3 Extraction du texte . . . 47
3.3.4 Localisation des personnages . . . 48
3.3.5 Vers une approche globale . . . 49
3.4 Point de vue représentation du domaine . . . 49
3.4.1 Des encodages pour le contenu . . . 49
3.4.2 A Comics Ontology . . . 51
3.5 Des approches hybrides . . . 52
3.5.1 Annotation consensuelle d'images . . . 53
3.5.2 Réduction du fossé sémantique . . . 55
3.5.3 Support pour l'analyse d'images . . . 60
3.6 Conclusion . . . 61

II Proposition d'un cadre pour l'analyse et l'interprétation automatisées d'images de bandes dessinées et son évaluation. 63

4 Présentation des ontologies proposées 65
4.1 Introduction . . . 65
4.2 Domaine de l'image . . . 66
4.2.1 Image et régions d'intérêt . . . 66
4.2.2 Extracteurs . . . 69
4.2.3 Vérité terrain . . . 71
4.2.4 Évaluation . . . 73
4.3 Domaine de la bande dessinée . . . 75
4.3.1 Des albums et des planches . . . 76
4.3.2 Contenu des planches . . . 77
4.3.3 Spécialisation du contenu . . . 81
4.4 Interaction entre les deux ontologies . . . 83
4.5 Conclusion . . . 83

5 Construction d'une vérité terrain 85
5.1 Introduction . . . 85
5.2 Description du contenu . . . 86
5.2.1 Les albums . . . 86
5.2.2 Les fichiers images . . . 86
5.2.3 Les pages . . . 87
5.2.4 Les cases . . . 87
5.2.5 Les phylactères . . . 87
5.2.6 Le texte . . . 88
5.3 Annotations visuelles . . . 88
5.3.1 Les cases . . . 88
5.3.2 Les bulles . . . 89
5.3.3 Les lignes de texte . . . 89
5.3.4 Les personnages . . . 91

5.4 Annotations sémantiques . . . 91
5.4.1 Les pages . . . 91
5.4.2 Les cases . . . 92
5.4.3 Les bulles . . . 92
5.4.4 Les lignes de texte . . . 93
5.5 Stockage et évaluation . . . 93
5.5.1 Structure de la vérité terrain . . . 93
5.5.2 Évaluation de l'homogénéité . . . 97
5.6 Conclusion . . . 98

III Vers une compréhension globale du contenu d'une bande dessinée. 101

6 Analyse contextuelle des planches 103
6.1 Introduction . . . 103
6.2 Description générale . . . 104
6.2.1 Communication entre les deux systèmes . . . 104
6.2.2 Contraintes pour l'extraction . . . 105
6.3 Boucle d'interaction . . . 107
6.3.1 Reconnaissance des éléments simples . . . 108
6.3.2 Reconnaissance des éléments complexes . . . 111
6.4 Évaluation des performances du système . . . 117
6.4.1 Protocole d'évaluation . . . 117
6.4.2 Évaluations individuelles . . . 119
6.4.3 Évaluation générale . . . 121
6.5 Conclusion . . . 122

7 Navigation et recherche d'informations 125
7.1 Introduction . . . 125
7.2 Génération de la séquence de lecture . . . 126
7.2.1 Problématique . . . 126
7.2.2 Relations spatiales . . . 126
7.2.3 Ordonnancement des cases . . . 129
7.2.4 Intégration à l'ontologie . . . 131
7.2.5 Évaluation . . . 133
7.3 Contribution pour la recherche d'informations . . . 135
7.3.1 Contexte . . . 135
7.3.2 Navigation au sein d'un treillis . . . 136
7.3.3 Optimisation du calcul des successeurs . . . 140
7.3.4 Application et évaluation . . . 142
7.4 Conclusion . . . 146

8 Conclusion générale **149**
8.1 Conclusion . . . 149
8.1.1 Deux ontologies et une base annotée... . . . 150
8.1.2 ...pour la compréhension d'images... . . . 151
8.1.3 ...et la recherche d'informations dans une BD. . . . 151
8.2 Perspectives . . . 151
8.2.1 Enrichir les interactions,... . . . 151
8.2.2 ...les capacités de raisonnement... . . . 152
8.2.3 ...et la recherche d'informations. . . . 152

Publications **153**

Bibliographie **155**

A Annexes **171**
A.1 Protocole de construction de la vérité terrain . . . 171
A.1.1 Pré-traitement . . . 171
A.1.2 Cases . . . 171
A.1.3 Zones de texte . . . 171
A.2 Rappel sur l'utilisation de Toonshop . . . 172
A.2.1 Créer une vérité terrain sur une page . . . 172
A.2.2 Créer un polygone . . . 172
A.2.3 Modifier un polygone . . . 172
A.2.4 Supprimer un polygone . . . 173
A.2.5 Utilisation du zoom et navigation . . . 173
A.2.6 Ajouter ou modifier des métadonnées . . . 173
A.3 Formalismes de bande dessinée . . . 173
A.3.1 ComicsML . . . 173
A.3.2 A Comics Ontology . . . 177

Table des figures

1.1 Illustration du style franco-belge. 3
1.2 Illustration du comics nord-américain. 4

2.1 Sources de données de DBpedia en 2007. 17
2.2 Sources de données de DBpedia en 2011. 18
2.3 Illustration d'un triplet RDF. 18

3.1 Étirement horizontal d'une case. 39
3.2 Étirement vertical d'une case. 40
3.3 Intégration des bulles. 42
3.4 Différents contours de bulles 42
3.5 Différentes formes de bulles 43
3.6 Bulle de pensée 43
3.7 Hors bulle 44
3.8 Les encadrés 44
3.9 Différentes polices de caractères 45
3.10 Les onomatopées 46
3.11 Différence entre bulle ouverte et bulle fermée 47
3.12 Exemple de déformation de personnage 48
3.13 Hétérogénéité sémantique 53
3.14 Interprétation d'une image 56
3.15 Fossé sémantique 57
3.16 Object Ontology de [Mezaris 2003] 57
3.17 Extraits de la hiérarchie de relations spatiales de [Hudelot 2008b] . . 59

4.1 Ontologie image - Première étape 69
4.2 Ontologie image - Deuxième étape 71
4.3 Ontologie image - Troisième étape 74
4.4 Ontologie image - Quatrième étape 76
4.5 Ontologie bande dessinée - Première étape 77
4.6 Ontologie bande dessinée - Deuxième étape 80
4.7 Propriétés du modèle de la bande dessinée 81
4.8 Spécialisation des concepts *Character*, *Balloon* et *TextLine* 82
4.9 Interaction entre les ontologies image et bande dessinée 84

5.1 Annotation des cases 89
5.2 Annotation des bulles de dialogues 90
5.3 Annotation des lignes de texte 90
5.4 Annotation des personnages 91
5.5 Les différents types de contour de bulle 93

5.6 Style d'affichage des éléments de la vérité terrain 94
5.7 Mesure d'homogénéité de la base eBDtheque 97
5.8 Page de mesure d'erreur 99

6.1 Représentation du système d'analyse 104
6.2 Boucle de traitement 105
6.3 Concordance entre notre conceptualisation et la base eBDtheque . . 107
6.4 Cases d'illustration 108
6.5 Boucle d'interaction - Étape 1.1 109
6.6 Boucle d'interaction - Étape 1.2 110
6.7 Boucle d'interaction - Étape 1.3 111
6.8 Exemple de flèche complexe 112
6.9 Localisation de la zone hypothétique du locuteur 114
6.10 Position estimée des régions contenant des personnages 115
6.11 Boucle d'interaction - Étape 2.1 115
6.12 Boucle d'interaction - Étape 2.2 116
6.13 Boucle d'interaction - Étape 2.3 117
6.14 Résultats globaux 122

7.1 Exemple de séparation de cases. 127
7.2 Les huit relations du RCC-8. 127
7.3 Différents cas de calcul d'orientation 129
7.4 Cas particulier produisant un ordre de lecture incohérent. 131
7.5 Relation de succession entre les cases 132
7.6 Exemple de tri d'un ensemble de cases 132
7.7 Exemple de mise en page extrême 134
7.8 Interface de requêtage par navigation de Sewelis 136
7.9 Case descriptive 144
7.10 Treillis d'exemple 145
7.11 Treillis d'exemple inversé 146

A.1 Hiérarchie de concepts de l'ontologie proposée par [Rissen 2012] . . . 177

Liste des tableaux

2.1 Onze axiomes tautologiques de la logique booléenne 25
2.2 Constructeurs composant les logiques de description 29
2.3 Constructeurs logiques pour les langages $\mathcal{AL}$, $\mathcal{ALE}$, $\mathcal{ALCN}$ et $\mathcal{SHOIN}$. 30
2.4 Sémantique des constructeurs des logiques de description 31

6.1 Coefficient de décalage de la zone en x et en y pour une orientation de la flèche donnée. 113
6.2 Évaluations indépendantes du *système haut niveau* 120

7.1 Résultats de notre méthode d'ordonnancement des cases 133
7.2 Une opération de ET logique permet de calculer l'extension d'un ensemble d'attributs 142
7.3 Contexte d'exemple 143
7.4 Temps de calcul (en secondes) des successeurs immédiats de $\bot$ et des prédécesseurs immédiats de $\top$. 144

A.1 Description des propriétés de l'ontologie proposée par Paul Rissen . . 178

Chapitre 1

Introduction générale

Sommaire

1.1 La bande dessinée,... 1
1.2 ...un patrimoine mondial... 2
1.2.1 La bande dessinée franco-belge 2
1.2.2 Le comics nord-américain 3
1.2.3 Le manga japonais 4
1.2.4 Le webcomic 5
1.3 ...en pleine mutation... 5
1.3.1 L'exemple des autres médias 5
1.3.2 Le cas de la bande dessinée 6
1.4 ...générant de nouveaux défis. 8

1.1 La bande dessinée,...

Tout le monde connaît la bande dessinée. Même les moins fervents de ses adeptes en ont au moins lu une, un jour, dans un album ou au bas de la dernière page d'un journal. Elle a longtemps souffert, et souffre parfois encore, d'une image de sous-culture mineure, produit hybride de la littérature et des arts de l'image incapable d'en retranscrire toute la noblesse. Ce manque de reconnaissance culturelle peut aujourd'hui encore être constaté au travers des médias généralistes, ayant plus facilement tendance à donner leur avis sur les sorties cinéma récentes que sur les dernières publications BD, ainsi qu'à travers le problème, soulevé par [Guilbert 2011], de la place accordée à ce média dans les musées.

La bande dessinée mériterait pourtant plus de crédit. Dès 1964, Claude Beylie n'hésite pas à proposer l'appellation de *neuvième art* [Beylie 1964], reprise peu après par [Morris 1964] et [Lacassin 1971]. Pourquoi la bande dessinée réduirait-elle nécessairement la valeur des arts desquels elle s'inspire ? D'après la théorie des arts de Ricciotto Canudo [Canudo 1923], la peinture et la sculpture sont dérivées de l'architecture, tandis que la musique a permis l'apparition de la danse et de la poésie. Ceux-ci sont-ils pour autant relégués au rang d'arts mineurs ou vus comme de vulgaires extensions d'arts qui seraient alors considérés comme plus nobles ? Ce raisonnement peut être transposé à la bande dessinée qui, comme le souligne [Groensteen 1996], a toutes les cartes en main pour être le berceau d'œuvres ayant le potentiel holistique de transcender leurs origines afin d'aboutir à un résultat à la fois plus riche et inédit.

La bande dessinée est un art séquentiel. Scott McCloud la définit comme « une suite d'images picturales et fixes volontairement juxtaposées en séquences » dans son ouvrage *L'Art Invisible* [McCloud 1993], universellement considéré comme référence en la matière. Il y dissèque les codes développés par les auteurs de bande dessinée au cours de l'évolution du média, et met en évidence ce que les mots et les images permettent d'exprimer lorsqu'ils sont combinés sur une planche.

Nous présentons dans cette introduction un rapide panorama des différents visages de la bande dessinée à travers le monde, ses origines et ses interrogations quant à la place qu'elle doit prendre dans un monde devenant de plus en plus dématérialisé. Les enjeux de cette transformation, ainsi que les problématiques scientifiques qu'elle soulève, sont au cœur des travaux présentés dans ce manuscrit.

1.2 ...un patrimoine mondial...

S'il existe une infinité de styles de bande dessinée différents, dépendant à la fois de l'auteur, de l'époque et de la culture, il est possible d'esquisser une première classification grossière du patrimoine mondial à travers quatre grandes catégories.

1.2.1 La bande dessinée franco-belge

La bande dessinée franco-belge sous sa forme moderne est, comme son nom l'indique, née dans l'ouest de l'Europe à la fin du XIX^{e} siècle. D'abord principalement publiée de manière épisodique dans des journaux spécialisés [François 2005], sa popularité grandissante en a fait un acteur de poids dans le monde de l'édition, représentant un dixième des livres vendus en 2012 [1]. Elle est caractérisée par un découpage en albums, d'un format légèrement supérieur à un A4, composés d'une quarantaine de pages environ (on parle plutôt de planches). Le contenu de ces planches est généralement linéaire. Les cases, en moyenne au nombre de douze, ont tendance à être clairement séparées les unes des autres. Le style a également pour caractéristique d'être très récitatif, avec un nombre important de bulles, contenant un grand volume de texte [Dogan 2013].

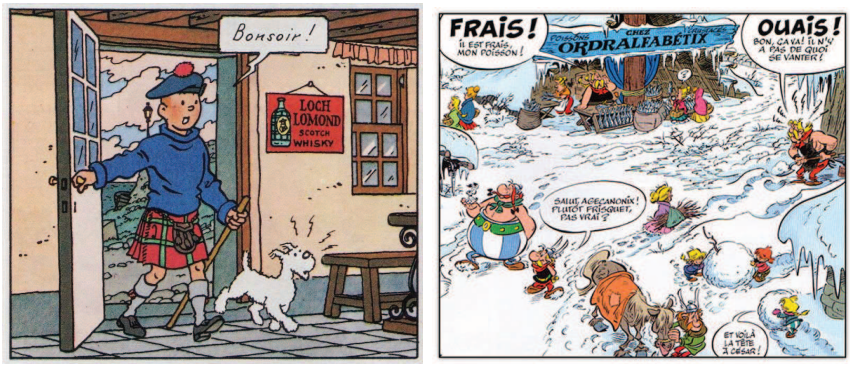
Le style franco-belge est souvent divisé en deux écoles [Deyzieux 2008]. L'école de Bruxelles (figure 1.1a), dite de la *ligne claire*, dont les représentants les plus célèbres sont Hergé (Tintin) ou encore Alain Saint-Ogan (Zig et Puce). Elle se distingue par un dessin extrêmement précis, mis exclusivement au service du récit, où la clarté prime sur tout le reste. Chaque forme est invariablement délimitée par un contour noir et qualifiée d'une couleur unie qui ne souffrira d'aucun dégradé et que l'on retrouvera tout au long de l'album.

A l'opposé, l'école dite de Marcinelle (figure 1.1b) adopte un style plus débridé, mettant l'accent sur la caricature et le comique plus que sur la narration, aboutissant à des albums moins verbeux que ceux de leurs confrères de Bruxelles. Le trait et le lettrage se veulent volontairement moins précis, plus libres. Les auteurs généralement

1. Source : Syndicat National de l'Edition

associés à ce style de BD sont Franquin (Spirou), Uderzo (Astérix) ou encore Morris (Lucky Luke).

Bien qu'ayant évolué de manière continue au cours de son histoire, notamment sur ses procédés narratifs ou sur les thèmes évoqués, le style franco-belge reste encore aujourd'hui globalement respectueux des bases fondamentales de mise en page introduites à ses origines.



(a) Bruxelles (b) Marcinelle

FIGURE 1.1 – Illustration du style franco-belge. (a) L'école de Bruxelles. Crédits : Hergé - L'île Noire - Casterman. (b) L'école de Marcinelle. Crédits : René Goscinny et Albert Uderzo - Astérix et Cléopâtre - Dargaud.

1.2.2 Le comics nord-américain

La bande dessinée nord-américaine trouve son origine au milieu du XIX^{e} siècle avec, comme en Europe, une première publication dans la presse. On parle alors de *comic strips*, littéralement *bandes comiques*, ou encore de *funnies* [Harvey 1994]. Les termes *comic books*, puis *comics*, n'arriveront que dans les années 1930, d'abord pour qualifier des recueils de comic strips [Decker 1983], puis pour désigner des albums mettant en scène des super-héros, alors tout juste naissants, qui rencontreront par la suite le succès que l'on connaît (1938 : Superman, 1939 : Batman, Flash, 1940 : Captain America, etc. [Coogan 2009]). Le format standard d'un comic book est légèrement plus petit que celui des BD européennes, environ dix-sept sur vingt-quatre centimètres, pour une cinquantaine de pages. Il est généralement recouvert d'une couverture souple en papier glacé.

Si la mise en page des premiers comics était similaire à celle employée par les auteurs européens, leurs confrères américains ont rapidement su s'affranchir des codes du genre afin de créer des planches au contenu moins régulier. Bien que l'on puisse également retrouver ce genre de structure dans les autres cultures, le comics moderne est souvent associé à une mise en page torturée, composée de cases abandonnant le format rectangulaire et supprimant la frontière les séparant de leurs voisines (voir exemple figure 1.2).

FIGURE 1.2 – Illustration du comics nord-américain. Crédits : John Ostrander et Jan Duursema - Star Wars Legacy #46 - Dark Horse Comics.

1.2.3 Le manga japonais

La bande dessinée japonaise a également vu le jour dans la deuxième moitié du XIX^{e} siècle. Le terme *manga*, signifiant littéralement « images dérisoires », a été introduit par Hokusai qui utilisait le terme pour désigner ses esquisses [Gabilliet 2009]. Le manga est très rapidement divisé en catégories distinctes, chacune visant un type de public bien défini, le *shonen* pour les adolescents, le *shojo* pour les jeunes filles, le *seinen* pour les adultes, etc. [Bouissou 2008]. Il adopte un format épisodique, publié de manière régulière dans des magazines. Ces épisodes sont ensuite réédités en volumes d'en moyenne deux cents pages à l'occasion d'une deuxième exploitation commerciale.

Le manga se distingue de la bande dessinée occidentale sur plusieurs points. Premièrement, les pages et les cases se parcourent de droite à gauche afin de respecter le sens de lecture de l'écriture japonaise. Deuxièmement, alors que la couleur a joué un rôle important dans le reste du monde [Gabilliet 2005], les mangas sont historiquement réalisés et publiés en noir et blanc pour des raisons économiques. Des jeux de trames sont alors utilisés pour simuler les ombres, les textures et les mouvements [Massé 2006]. Le dessin prend une place importante dans la narration, le *mangaka* (dessinateur de mangas) ayant à sa disposition toute une palette de codes graphiques incrustée dans l'inconscient collectif des lecteurs japonais, telles que la « goutte de tension » ou la « veine de colère » [Paez 2008].

L'organisation des cases est souvent plus irrégulière encore que celle des comics modernes et met en avant les personnages aux dépends du décor.

1.2.4 Le webcomic

Composant notre quatrième catégorie, le terme « webcomic » désigne avant tout une bande dessinée originellement publiée sur Internet et n'entretient pas de rapport particulier avec le *comics* américain. Les premiers webcomics disponibles en ligne datent de la fin des années 1980, début des années 90, c'est à dire à l'époque même de l'invention du World Wide Web, mais leur véritable envol se fera une décennie plus tard [Garrity 2011]. La multiplication des plateformes de blogging et des réseaux sociaux ouvre alors les portes de l'auto-édition à tout un chacun. Des plateformes de publication, telles que *webcomics.fr* voient le jour, fournissant des outils pour renforcer l'interactivité entre les auteurs et les lecteurs [Zouari 2009]. L'absence de support physique libère les auteurs de toutes les contraintes matérielles apportées par l'impression. Au contraire, le support numérique et les technologies du web leur offrent des perspectives de mise en page et d'interaction inédites [Chiron 2008]. Le webcomic est le premier pas de la bande dessinée dans l'ère du numérique, précurseur d'une profonde remise en question de la nature du média face aux nouvelles technologies.

1.3 ...en pleine mutation...

1.3.1 L'exemple des autres médias

Nous l'avons vu, la bande dessinée représente une part importante de l'héritage culturel de nombreux pays, notamment en Europe de l'ouest, au Japon et aux États-Unis. Elle n'a cependant pas encore bénéficié du même niveau d'attention que la musique, le cinéma ou même, dans une moindre mesure, la littérature dite classique concernant son adaptation au format numérique. Les services se sont en effet multipliés pour les deux premiers, que ce soit au niveau des plateformes de diffusion musicale [ODLM 2014], ou des fournisseurs de vidéo à la demande avec (SVOD) ou sans abonnement (VOD) [CCI 2008]. Malgré la prise de conscience tardive des enjeux du numérique par ces secteurs culturels, et même si du chemin reste à parcourir, notamment au niveau de la stabilité des modèles économiques mis en place [Generator Research 2013], l'offre, d'un point de vue technique, peut être considérée comme satisfaisante. Concernant la littérature, le marché du livre numérique reste pour l'instant confidentiel, ou tout du moins émergeant. Moins d'un lecteur sur cinq se déclare en effet intéressé par la lecture d'une œuvre dématérialisée, d'après les conclusions d'une étude Ipsos commanditée en 2010 par le Centre National du Livre [CNL 2010]. Si, une fois de plus, l'offre et le modèle économique choisis par les fournisseurs de contenu sont pointés du doigt pour expliquer ce relatif manque d'engouement [Benhamou 2011], certains aspects techniques sont également mis en cause. Notamment le caractère fastidieux que revêtent certaines actions, telles que le feuilletage ou l'annotation manuelle, manipulations naturelles sur le support papier originel. Cette dichotomie entre le *media*, l'œuvre en elle-même, et le *médium*, le support de transmission de l'œuvre, se retrouve, à un degré encore supérieur, dans

le monde de la bande dessinée.

1.3.2 Le cas de la bande dessinée

Le processus de conversion et d'adaptation au numérique d'une bande dessinée imprimée n'est pas aussi simple que celui mis en place pour la dématérialisation d'un film, d'un album de musique ou d'un roman. En effet, la totalité d'un film peut être réduite à une suite d'images synchronisées sur une bande son. Visionner ces images dans le bon ordre et à la bonne cadence permet de retrouver le contenu initial, quel que soit le support. Dans le même esprit, un roman n'est en définitive qu'une suite de mots. Même si l'on a vu que l'expérience de lecture peut être altérée par le changement de médium, lire ces mots, dans le bon ordre, sur du papier ou sur un écran ne change rien, ni au contenu, ni à la dimension artistique de l'œuvre. La bande dessinée diffère de ces derniers par le fait que le média est en lui-même intimement lié au médium de diffusion utilisé. Elle s'éloigne notamment du cinéma, auquel elle emprunte beaucoup de codes, sur la notion de *juxtaposition* des images. Là où, au cinéma, les images sont toutes de taille identique et où chacune vient remplacer la précédente sur la totalité de l'espace de visualisation, les cases d'une bande dessinée sont de tailles variables et organisées spatialement sur cet espace. À ces deux particularités s'ajoute le fait que le lecteur a la possibilité de parcourir d'un coup d'œil toutes les cases d'une même page et n'est donc pas physiquement contraint de parcourir l'histoire dans un ordre prédéfini. Ces spécificités sont autant d'outils au service de l'auteur lui permettant de mettre en scène son histoire. Dès lors, changer le support de l'œuvre, la surface de lecture ou la manière de séquencer les cases, implique une modification de la mise en scène qui peut, dans le pire des cas, être néfaste au récit. Tout l'enjeu de la bande dessinée numérique consiste donc à trouver comment tirer parti des spécificités de ces nouveaux supports que sont les ordinateurs, Internet, les tablettes et autres smartphones.

Comme le souligne Gilles Ratier, secrétaire général de l'Association des Critiques et journalistes de Bande Dessinée (ACBD), dans le bilan 2013 de l'association [Ratier 2013], la sémantique du terme *bande dessinée numérique* reste encore imprécise. Celui-ci est en effet couramment utilisé pour désigner un large ensemble de contenus hétérogènes, allant de la simple numérisation d'œuvres imprimées aux limites du turbo-média (les récits en images interactifs), en passant par les webcomics et la réalité augmentée qui permet d'étendre l'univers de l'œuvre grâce à l'utilisation de la technologie. De nombreux débats sont organisés pour essayer de définir la place que peut et que doit prendre la bande dessinée dans l'univers numérique ; de manière récurrente au *Festival international de la bande dessinée* d'Angoulême, ou plus ponctuelle comme la conférence *La bande dessinée : du papier au numérique* organisée en février 2012 à la médiathèque de la Rochelle dans le cadre d'une ANR sur l'étude des impacts du numérique sur la culture [Collectif PANIC 2011].

Depuis 2008, de nombreux acteurs se sont lancés dans l'aventure de l'édition ou de la distribution de BD numériques, proposant différents services et différentes

expériences de lecture. Plusieurs plateformes de diffusion de bandes dessinées numérisées ont vu le jour, telles que Numilog [2], DigiBDi [3] ou encore Izneo [4]. Grâce à des partenariats conclus avec les poids lourds de l'édition française que sont, entre autres, Hachette, Dargaud, Casterman, ou encore Dupuis, ainsi qu'à la multiplication des acteurs, une grande majorité de l'offre imprimée est maintenant accessible en numérique. L'américain ComiXology [5], leader incontesté aux Etats-Unis, a vu son arrivée en France naître de son association avec la maison Delcourt en 2013. Outre la possibilité offerte d'acheter les albums numérisés à l'unité, ces plateformes tentent différentes approches économiques avec, par exemple, la mise en avant de liens vers l'achat des œuvres physiques chez DigiBDi ou la proposition d'un abonnement mensuel de « bande dessinée à la demande » pour Izneo. La scène indépendante n'est pas en reste avec en premier lieu les blogs BD individuels, mais également les plateformes de centralisation des productions indépendantes, telles que webcomics.fr [Zouari 2009] ou EspritBD [6] qui sont souvent le théâtre de la percée d'usages créatifs inédits, rendus possibles sur ces nouveaux supports numériques.

D'autres acteurs se concentrent tout particulièrement sur les plateformes mobiles qui leur offrent la possibilité d'introduire de nouvelles habitudes et de nouveaux réflexes, plus naturels, de lecture numérique. Si outre-atlantique, la maison d'édition Marvel [7] s'est clairement positionnée comme un acteur d'envergure du trans-média, étendant l'univers de ses comics sur d'autres supports via notamment des applications de réalité augmentée (sans parler des films), les acteurs de l'hexagone ne sont pas en reste. AveComics [8], créée par la société montpelliéraine Aquafadas, a par exemple été la première plateforme à proposer, dès 2009 (une époque où l'iPad n'existait pas encore), des outils permettant de guider la lecture d'une page de bande dessinée par des effets de zoom et de transition sur et entre des zones de l'image, afin de rendre la lecture possible sur des petits écrans. D'autres initiatives, telles que Sequencity [9], production récente de la société Actialuna, se concentrent plus particulièrement sur l'ergonomie de lecture et cherchent à transposer sur tablette l'expérience offerte en librairie en permettant par exemple le feuilletage rapide et intégral d'une œuvre et en intégrant des moyens de communications avec des professionnels à même de fournir des recommandations personnalisées au lecteur.

Mais la création d'expériences inédites et la transformation du patrimoine existant ne sont pas les seuls enjeux portés par cette révolution numérique. L'indexation et la valorisation de ce patrimoine, ainsi que la préservation d'œuvres uniques, rares et fragiles sont également au cœur des préoccupations des libraires, archivistes et

2. www.numilog.com
3. www.digibidi.com
4. www.izneo.com
5. www.comixology.com
6. www.espritbd.fr
7. www.marvel.com/comics
8. www.avecomics.com
9. www.sequencity.com

collectionneurs. La recherche d'information au sein de bases de planches numérisées est une tâche qui n'est pas triviale. L'annotation manuelle du contenu est lourde et complexe, particulièrement lorsque l'on souhaite pouvoir effectuer des requêtes complexes telles que « les cases des albums de Tintin datant d'avant 1962 où apparaissent Milou et le Capitaine Haddock et où celui-ci crie les mots "mille milliards de mille sabords" ». L'apport d'une réponse à ce genre de requêtes nécessite aujourd'hui, de la part des archivistes, une connaissance aiguë des albums, ne les dispensant toutefois pas d'une phase de recherche manuelle chronophage et fastidieuse.

Certains s'intéressent également à l'étude de l'évolution des œuvres d'une collection, du style d'un auteur ou d'un mouvement particulier à travers le temps. C'est notamment le cas de [Pissavy-Yvernault 2013] détaillant et analysant l'évolution du dessin de Spirou à travers ses soixante-quinze ans d'existence.

1.4 ...générant de nouveaux défis.

La bande dessinée est un art majeur, chargé d'histoire, mêlant images et textes de manière unique afin de transmettre un message au lecteur. Elle possède ses codes graphiques qui, s'ils sont pour la plupart devenus consensuels au fil du temps, sont en constante évolution et sont régulièrement mis à mal par des auteurs expérimentant de nouveaux vecteurs narratifs pour mettre en scène leurs histoires. Les nouveaux usages introduits par l'ère numérique, que l'on parle de transformation de mise en scène pour une lecture sur de nouveaux médiums, d'enrichissement par l'analyse du contenu ou encore de recherche d'information, se heurtent, une fois appliqués à grande échelle, à plusieurs verrous scientifiques. Ceux-ci sont liés, d'un côté, à l'extraction des éléments constitutifs de l'image et, d'un autre côté, à l'organisation de ces composants dans une structure permettant d'en faire émerger une sémantique. Le projet eBDtheque [10] a été lancé en 2011 à l'université de la Rochelle dans le but de proposer des solutions à ces différents problèmes. Les travaux ont été réalisés dans le cadre de deux thèses, menées en parallèle. Celle de Christophe Rigaud [Rigaud 2014a], portant sur le traitement des images de bande dessinée numérisées, et celle, présentée dans ce document, s'attachant à la création d'un cadre d'analyse et de raisonnement. Cette collaboration a donné lieu à des rapprochements entre les deux thèses qui seront mis en avant dans plusieurs sections.

La première partie de ce manuscrit présente les ressources, les outils et les connaissances dont nous disposons pour conduire l'analyse d'une planche de bande dessinée. La compréhension sémantique du contenu d'une page, si elle est naturelle pour un être humain, ne l'est pas pour une machine. Aussi, si l'on cherche à caractériser celui-ci de manière automatique, il nous faut le formaliser de façon à le rendre compréhensible par un ordinateur. Le domaine de recherche très actif de la *représentation des connaissances* a produit des outils, notamment les ontologies, permettant d'aborder ce genre de problème. Le chapitre 2 présente une synthèse de

10. l3i.univ-larochelle.fr/eBDtheque

leurs origines, leur composition et leurs mécanismes de fonctionnement. En amont de cela, il est nécessaire de pouvoir détecter et reconnaître les composantes visuelles d'une planche de bande dessinée, c'est à dire être capable d'estimer qu'une case se situe à tel endroit ou qu'un personnage se situe à tel autre. C'est une tâche qui n'est pas triviale et les techniques utilisées communément en analyse d'images, ou plus précisément dans notre cas en analyse d'images de documents, ne produisent jamais, quelles que soient leur complexité, de résultats parfaits en toutes circonstances. Elles peuvent néanmoins s'appuyer sur une étude des codes graphiques adoptés par les auteurs de bande dessinée afin d'adapter leurs traitements. Nous ferons une synthèse de ces codes au cours du chapitre 3 puis nous passerons en revue les travaux s'étant attaqués à l'analyse visuelle d'images de bandes dessinées. Nous verrons également les différentes façons dont les technologies de représentation de connaissances peuvent être mises à contribution pour l'analyse et l'annotation d'images.

La deuxième partie de ce document est consacrée à la présentation d'un cadre pour l'analyse et l'interprétation d'images de bande dessinée. Ce cadre est composé, d'une part, d'une ontologie formalisant les notions relatives au domaine de l'analyse d'images et, d'autre part, d'une ontologie formalisant les code graphiques de la bande dessinée. Des équivalences entre leurs concepts respectifs permettent de lier ces deux ontologies et de les utiliser conjointement pour différentes tâches. Leur conceptualisation et leur formalisation sont détaillées au sein du chapitre 4. Nous présentons ensuite, au cours du chapitre 5, le protocole de construction et le contenu d'un jeu de données de vérité terrain, développé pour permettre l'évaluation des travaux d'analyse visuelle et sémantique de planches de BD. Celui-ci a été publié et rendu publiquement accessible à la communauté scientifique qui a d'ores et déjà commencé à l'utiliser.

Enfin, la troisième et dernière partie présente les différents contextes dans lesquels le cadre que nous avons développé a pu être utilisé. Le chapitre 6 présente la mise en place d'un framework complet pour l'analyse de planches, principal fruit de la collaboration entre cette thèse et celle de Christophe Rigaud. A travers une boucle d'interaction entre le système d'analyse d'images et le système de raisonnement, une compréhension globale et formelle du contenu de la planche analysée est acquise de manière itérative. Les résultats de ces travaux sont évalués grâce à la vérité terrain présentée dans la partie deux. Pour finir, le chapitre 7 contient la présentation de plusieurs solutions proposées pour la navigation au sein de bases de bandes dessinées numérisées et annotées. Nous présentons tout d'abord un mécanisme de production de la séquence de lecture d'une planche, en vue d'un transfert de l'œuvre vers de nouveaux supports, basé sur l'exploitation des relations spatiales entre les composants graphiques de l'image. Notre méthode présente la particularité de tenir compte de plusieurs spécificités de mise en page, laissant dans certains cas le lecteur maître de son parcours des cases, comme il peut l'être à la lecture d'une œuvre imprimée. Nous présentons ensuite des travaux effectués dans le but d'optimiser la navigation au sein d'un treillis de concepts pouvant accueillir les données

de notre modèle, facilitant ainsi la recherche d'information au sein de celui-ci.

Première partie

Des connaissances et des outils pour une analyse visuelle et sémantique de la bande dessinée.

Chapitre 2

Représentation des connaissances

Sommaire

2.1 Introduction 13
2.2 L'exemple du web sémantique 14
2.2.1 Du web des données au web social 14
2.2.2 Le web sémantique 15
2.3 Les ontologies 19
2.3.1 Définition 19
2.3.2 Composition 20
2.4 Les Logiques de Description (LD) 23
2.4.1 Introduction 23
2.4.2 Différentes logiques fondatrices 23
2.4.3 Une famille de logiques 28
2.4.4 Sémantique 31
2.4.5 Raisonnement 31
2.5 Conclusion 33

2.1 Introduction

La représentation des connaissances est l'une des disciplines fondamentales couvertes par l'intelligence artificielle. Mais qu'est-ce que la connaissance et pourquoi vouloir la représenter ? La connaissance que nous avons de certains domaines, nous humains, est ce qui nous permet de déduire des informations à partir des données que nous recevons. Nous sommes par exemple le 15 décembre, je déduis de ma connaissance du calendrier et des fêtes chrétiennes qu'il ne reste que 10 jours avant Noël. A partir de ces informations nous pouvons prendre des décisions et choisir d'effectuer certaines actions, comme par exemple se dépêcher d'aller acheter des cadeaux.

Dans un monde devenu ultra-informatisé et où les données circulent sur les réseaux à flots continus, l'interprétation automatique de ces données est devenue un enjeu de premier ordre. L'objectif poursuivi par la communauté scientifique en représentation des connaissances est donc la mise en place de systèmes permettant cette interprétation des données.

L'un des exemples d'application les plus percutant, introduit il y a une dizaine d'années, est sans doute celui du web sémantique.

2.2 L'exemple du web sémantique

En ce début du vingt-et-unième siècle, nous avons été les témoins et les acteurs de l'expansion formidable du réseau Internet et de la place qu'il prend dans nos vies personnelles et professionnelles. Passé du statut de curiosité globalement méconnue du grand public dans les années quatre-vingt dix, le réseau s'est, petit à petit, rendu littéralement indispensable dans le quotidien du plus grand nombre. Parmi tous les services ayant émergé au cours de son développement, le web est sans conteste l'un des plus populaires et des plus utilisés quotidiennement.

2.2.1 Du web des données au web social

La première version du web, telle que présentée par [Berners-Lee 1991] et aujourd'hui rétrospectivement connue sous le nom de *web 1.0*, a pris son envol dans les années 1990. La toile était alors constituée de pages web au contenu statique. L'information disponible en ligne circulait vers les utilisateurs de manière uni-directionnelle, les moyens de créer du contenu sur Internet étant alors trop complexes pour solliciter le visiteur, et a fortiori pour que celui-ci décide spontanément d'apporter sa pierre à l'édifice. C'est l'époque des premiers sites de e-commerce ainsi que des premiers moteurs de recherche. Si le réseau était déjà un support de communication, à la fois des entreprises aux particuliers via les sites web, mais également entre particuliers, notamment grâce aux courriels, il restait cependant confiné à une utilisation dans ce domaine pouvant être qualifiée de marginale.

L'émergence au début des années 2000 du web social a profondément modifié le rapport de l'utilisateur à Internet. Naturelle évolution du web connu jusqu'alors, [O'Reilly 2007] le définit comme une intelligence collective, un réseau non plus composé de logiciels mais de services s'améliorant lorsque les internautes les utilisent. Le changement de paradigme lui semble si important qu'il propose de parler de web 2.0, l'incrémentation matérialisant de manière symbolique le décalage s'étant opéré dans la manière d'interagir avec le réseau. Alors que ses utilisateurs étaient avant tout les consommateurs du contenu disponible en ligne, ils deviennent maintenant, petit à petit, également des créateurs. Les ressources humaines intellectuelles à même d'apporter leur contribution aux informations disponibles en ligne se multiplient, faisant mécaniquement augmenter de manière sensible le volume de données du web. Ce phénomène a véritablement commencé à prendre de l'ampleur lors de l'apparition des blogs personnels. Les internautes ont alors eu en mains, pour la première fois, les clés leur permettant de s'adresser au monde entier, que cela soit pour partager leurs photos de vacances ou leur analyse sur le dernier sujet de société en date. L'encyclopédie collaborative Wikipedia constitue également un exemple percutant de la puissance du web social. Celle-ci revendique, au moment de l'écriture de ces lignes, plus de 4,5 millions d'articles dans sa version anglaise [1], contre 85000 pour

1. en.wikipedia.org/wiki/Wikipedia :Statistics

une grande encyclopédie comme Britannica[2], avec un taux d'erreurs, recensé par [Giles 2005], tout à fait comparable entre les deux. Sans la contribution collective et active de ses utilisateurs, Wikipédia n'aurait probablement jamais pu espérer atteindre une couverture aussi vaste de la connaissance humaine.

Enfin, le web 2.0 a littéralement explosé avec l'arrivée des réseaux sociaux permettant le partage massif de contenus multimédias, tels que YouTube, Flickr ou Instagram, le relai d'actualités via Twitter ou Facebook, ou bien encore le partage d'informations localisées sur des lieux, des commerces ou des événements sur des services de localisation tels que Foursquare et Google Maps. Cette contribution individuelle apportée par tout un chacun à l'augmentation du volume d'information disponible en ligne a grandement été favorisée par, dans un premier temps, l'augmentation des débits, de la couverture et de la qualité de connexion à Internet puis, dans un second temps par la démocratisation des smartphones et autres dispositifs nomades auprès du grand public. Autrefois limitée à la mise à disposition de contenu textuel éventuellement illustré, la toile est devenue le théâtre d'un trafic phénoménal de contenus hétérogènes (plus de 32000 pétaoctets en 2012[3]), publiés par des milliards d'internautes à travers le monde.

2.2.2 Le web sémantique

Si l'augmentation de la taille du web est synonyme d'une source d'informations devenant jour après jour de plus en plus importante, ces informations sont toutefois publiées de manière arbitraire, sans contexte formel permettant d'identifier automatiquement et sans ambiguïté ce à quoi elles font référence.

Si l'on se rend sur la page Wikipédia de Barack Obama et que l'on y lit la première phrase *« Barack Hussein Obama II, né le 4 août 1961 à Honolulu dans l'État d'Hawaï, est un homme d'État américain. »* que peut-on instantanément en tirer comme information ? A la lecture de cette phrase, on comprend par exemple que son deuxième prénom est Hussein, le II nous indique qu'il n'est pas le premier de sa famille à s'appeler ainsi, qu'il aura bientôt 54 ans, qu'il est originaire d'une île de l'océan Pacifique et qu'il est maintenant politicien. Nous sommes capables de déduire tout cela de ce qui n'est en définitive qu'une suite de lettres car nous possédons une certaine *connaissance* des différents domaines impliqués dans notre phase d'interprétation. Nous avons premièrement une connaissance de la langue dans laquelle est écrite la phrase, une connaissance de son vocabulaire et de sa grammaire. Je serais personnellement bien en peine de déduire quoi que ce soit de la même phrase rédigée en gallois : *« Barack Hussein Obama II, a aned 4 Awst, 1961 yn Honolulu yn nhalaith Hawaii yn gwladweinydd Americanaidd »*. Cette connaissance de la langue nous indique notamment que lorsque l'on rencontre la locution *« né le »* suivie d'une date, celle-ci porte l'information du jour de la naissance de l'individu sujet du verbe *naître*. Deuxièmement, notre connaissance de l'être humain et des règles

2. www.britannica.com
3. www.cisco.com/go/vni

de société nous permet de déduire des informations de son nom complet (second prénom, deuxième du nom) et de sa date de naissance (son âge actuel), de son lieu de naissance et de sa profession. Troisièmement, notre connaissance de la géographie nous indique que l'état de Hawai est une île et que celle-ci est située dans l'océan Pacifique. La possession de connaissances dans d'autres domaines particuliers nous permettrait naturellement d'effectuer des déductions complémentaires.

Nous effectuons naturellement des liens entre les domaines de connaissances distincts que nous possédons (linguistique, êtres humains et géographie dans notre exemple) afin de conduire nos déductions. Nous savons par exemple que l'information temporelle suivant *« né le »* correspond à la date de naissance de la personne et que le terme suivant *« à »* est à la fois le lieu de naissance de l'individu et une entité géographique possédant des caractéristiques particulières. Nous sommes capable d'effectuer ces déductions car nous comprenons la sémantique des notions impliquées par la phrase, nous pouvons interpréter leur sens et les replacer dans un cadre plus large où des liens entre les domaines de connaissances distincts auxquelles elles appartiennent peuvent s'effectuer. Nous serions incapables de faire de même pour des domaines que nous ne comprenons pas, ou impliquant des notions que nous ne maîtrisons pas. C'est par exemple mon cas pour cette citation d'un article tiré d'une revue de neurochirurgie : *« Deux zones de conflit potentiel sont apparues : la pince entre le ligament sacro-tubéral et le ligament sacro-épineux, et le canal d'Alcock. Par ailleurs, le nerf peut être sténosé par le processus falciforme du ligament sacro-tubéral »*. Cependant, d'une phrase à partir de laquelle un humain peut, en fonction de ses connaissances, déduire un certain nombre d'informations, une machine n'y verra, dans le meilleur des cas, qu'une suite de lettres ou, dans le pire des cas, qu'une suite de valeurs booléennes.

Le concept de web sémantique, ou *web 3.0*, a été introduit par [Berners-Lee 2001] dans l'objectif de transférer cette compétence humaine aux systèmes informatiques. Le but est de rendre ces systèmes capables de comprendre, sinon d'interpréter, les données du web, ainsi que de pouvoir communiquer entre eux sur la base de références communes. Le web sémantique, comme ses prédécesseurs et contrairement aux bases de données relationnelles, n'est pas centralisé. Chaque utilisateur du réseau est libre de créer ses ressources (documents, pages web, contenu multimédia, mots...) et de faire référence aux ressources créées par d'autres. Celles-ci sont définies par des identifiants uniques sur le web, appelés URI (pour *Universal Resource Identifier*). A chaque ressource est associée une description formelle permettant de l'associer à un contexte, d'identifier le format dans laquelle elle est exprimée et de déterminer les relations qu'elle peut entretenir avec d'autres types de ressources. L'ensemble de ces ressources, associées à des règles de déduction basées sur la logique, telles que la subsomption (le terme *humain* est une spécialisation du terme *être vivant*), et l'instanciation (Barack Obama est un humain, il est donc également un être vivant), permet à un système informatique d'effectuer des raisonnements sur les ressources qu'il manipule afin d'en déduire des informations qui n'étaient pas explicitement exprimées en premier lieu.

Les descriptions associées aux ressources en définissent la *sémantique* de telle sorte qu'une machine puisse en comprendre la signification, ou tout du moins donner l'illusion de sa compréhension lors de ses interactions avec un utilisateur. Nous sommes en effet ici en présence du modèle de la chambre chinoise imaginé par [Searle 1980]. Peut-on considérer qu'un système qui manipule des ressources selon des règles explicitement définies, produisant des résultats logiquement déterminés, comprend réellement le sens de ses actions ? La compréhension d'un concept, telle qu'elle est considérée en logique, repose sur la connaissance de ce qui caractérise ce concept. Les systèmes du web sémantique manipulent des ressources sur lesquelles ils ont des connaissances, explicites et implicites, leur permettant d'effectuer des actions. En ce sens, il est donc possible de considérer qu'un tel système *comprend*, au sens logique du terme, la sémantique associée à ces ressources, à la mesure des descriptions leur étant associées.

De nombreuses initiatives ont vu le jour, à la fois pour venir peupler ce web sémantique de nouveaux contenus structurés, mais également pour apposer un cadre formel au contenu déjà en ligne. La plateforme DBpedia [Auer 2007] a par exemple à la fois pour ambition de structurer le contenu publié sur Wikipédia et de servir de noeud central pour la jonction d'ontologies à travers le web. Elle a connu une adoption rapide depuis sa mise en service en 2007 lorsqu'elle ne recensait que quelques millions de triplets (voir figure 2.1), jusqu'à aujourd'hui où elle en gère plusieurs milliards [Bizer 2009] (voir figure 2.2).

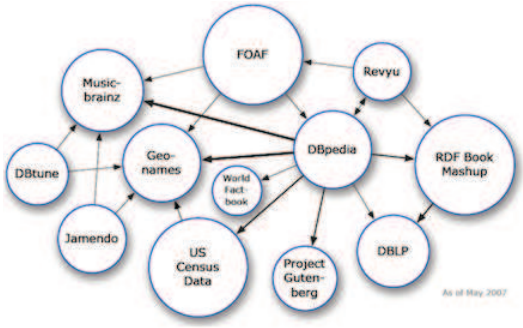


FIGURE 2.1 – Sources de données de DBpedia en 2007.

Cette description formelle des ressources du web sémantique s'organise autour de standards définis par le W3C. Le format RDF (pour *Resource Description Framework*) [Klyne 2004], développé Brickley et Guha, est devenu un standard en 1999 [Brickley 1999]. C'est un langage à balises basé sur XML (pour *eXtensible Markup Language*) [Bray 1998], fournissant un ensemble minimal d'opérateurs standards permettant de représenter des ressources ainsi que les relations entre elles sous la forme d'URIs. Il est, par définition, extensible via l'intégration de nouvelles notions, décrites par des URIs. Chaque déclaration de ressource, chaque relation entres deux

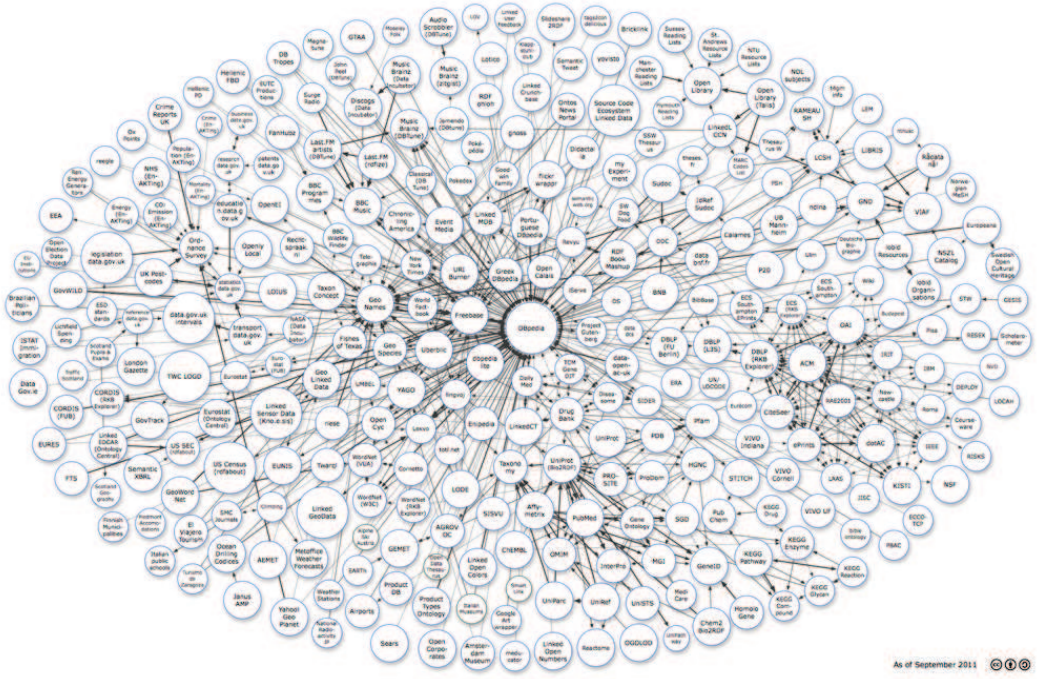


FIGURE 2.2 – Sources de données de DBpedia en 2011.

ressources, est représentée sous la forme d'un *triplet*. Un triplet est composé de trois entités, le sujet, la propriété et l'objet, et peut être représenté sous la forme d'un graphe orienté où les noeuds sont les ressources, tandis que les arcs sont les propriétés liant ces ressources, comme l'illustre la figure 2.3.

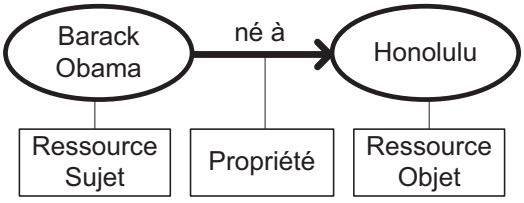


FIGURE 2.3 – Illustration d'un triplet RDF.

La définition formelle, mais utilisant des URIs factices, en syntaxe XML du triplet illustré en figure 2.3 est donné dans l'encadré 2.1.

Le langage a, peu à peu, été enrichi afin de permettre à la fois une plus grande expressivité dans la description des ressources du web et la normalisation de l'expression de certains types de relations usuelles. RDFS (pour *Resource Description Framework Schema*) a été proposé en 2000 [Brickley 2000] avant de devenir une recommandation du W3C en 2004. Il introduit entre autres la notion de *classes*, des

ensembles de ressources partageant des caractéristiques communes, ainsi que celles de *domain* et de *range*. Ces deux dernières permettent respectivement de contraindre le domaine de définition, le type du sujet, et l'ensemble d'arrivée, le type de l'objet, d'une propriété.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:exemple="http://www.exemple.org#">
  <rdf:Description rdf:about="http://www.exemple.org#BarackObama">
    <exemple:neA rdf:ressource="http://www.exemple.org#Honolulu"/>
  </rdf:Description>
</rdf:RDF>
```

ENCADRE 2.1 – Exemple de définition d'un triplet RDF.

Cette première brique de vocabulaire permet déjà de décrire un nombre conséquent de ressources du web. L'important volume de triplets générés est stocké dans des entrepôts de triplets (ou *triplestores*), conçus afin d'optimiser l'exécution de requêtes sur ces grandes bases de connaissance [Rohloff 2007]. Les requêtes sont exprimées en SPARQL [Prud'Hommeaux 2008] (*SPARQL Protocol And RDF Query Language*), un langage inspiré de la syntaxe de SQL [Date 1987] (*Structured Query Language*), largement utilisé pour les bases de données relationnelles.

Le vocabulaire de RDFS sera intégralement repris et enrichi par le langage OWL (*Web Ontology Language*) [McGuinness 2004], spécification du W3C et recommandation pour la construction d'ontologies. [Shadbolt 2006] souligne par ailleurs l'importance et le rôle prépondérant qu'elles doivent occuper dans la mise en œuvre du web sémantique, notamment grâce à leur capacité à représenter des systèmes complexes. Basées sur des paradigmes logiques, elles présentent en outre la possibilité d'intégrer des moteurs d'inférence permettant la déduction de connaissance implicite.

2.3 Les ontologies

2.3.1 Définition

Le terme *ontologie*, indénombrable, a été emprunté à la branche philosophique définie par Aristote comme « l'étude de l'être en tant qu'être », c'est à dire l'étude des propriétés caractéristiques de tout ce qui est. En informatique, *une* ontologie désigne un ensemble de définitions et de règles permettant de décrire la connaissance d'un domaine particulier, comme par exemple l'ornithologie, la linguistique, un hôpital ou encore la bande dessinée. Apparues à la fin du vingtième siècle, les ontologies ont très rapidement été perçues comme une technologie prometteuse, notamment grâce à leur capacité à faciliter les communications entre organisations et entre systèmes informatiques, ainsi que sur l'apport qu'elles peuvent fournir au développement et au maintien de suites logicielles [Uschold 1996].

La définition consensuelle de la notion d'ontologie, au sens de l'objet informatique, a évolué au cours du temps. Dès 1993, [Gruber 1993] la définit comme une « une spécification explicite d'une conceptualisation ». Une conceptualisation est une représentation abstraite et simplifiée du monde, d'un domaine ou d'un phénomène que l'on souhaite modéliser dans un but précis. Elle consiste en l'identification des concepts et des relations pertinentes pour le domaine étudié, accompagnée d'une définition précise et non-ambiguë pour chacun d'entre eux [Uschold 1995]. Une conceptualisation reste néanmoins dépendante de la capacité du concepteur à appréhender le domaine qu'il souhaite représenter, ainsi que de l'objectif final de cette modélisation. Des concepts et des relations possibles pour notre exemple sur Barack Obama seraient par exemple *Personne* et *Lieu*, ou encore *né-à*. Le choix du vocabulaire employé a de l'importance afin de limiter le risque de fausses interprétations quant à la signification attribuée à ces concepts et ces relations.

Une ontologie selon la définition de Gruber se doit donc de spécifier une telle conceptualisation, en définissant de manière explicite les concepts la composant.

Quelques années plus tard, [Borst 1997] complète cette définition en évoquant alors une « spécification formelle d'une conceptualisation partagée ». Une ontologie se doit donc, dans la mesure du possible, de spécifier une conceptualisation commune, consensuelle, plutôt que l'interprétation personnelle du concepteur. C'est particulièrement le cas pour les ontologies de haut niveau (*top-level ontologies* ou encore *foundational ontologies* dans la langue de Shakespeare) ayant un large champ d'action et destinées à être utilisées dans différents contextes [Schneider 2003]. Deuxièmement, cette spécification doit être formalisée dans un langage pouvant être interprété par une machine. Cela exclut de fait le langage naturel mais RDF, alors naissant, s'y prête particulièrement bien tout comme, naturellement, OWL qui deviendra par la suite le langage standard pour la création d'ontologies. Ce point est particulièrement important car il permet de s'assurer qu'un concept ou qu'une relation formalisée soit interprétée correctement, en accord avec la conceptualisation initiale, par tout un chacun. Cette formalisation permet par exemple d'éviter qu'une personne ne comprenant pas le français confonde la relation *né-à*, indiquant un lieu, avec une potentielle relation *né-en*, indiquant une date.

Ces deux définitions seront unifiées en 1998 par [Studer 1998] qui en proposera la synthèse qui semble aujourd'hui faire autorité [Guarino 2009] : *« Une ontologie est une spécification formelle et explicite d'une conceptualisation partagée. »*

2.3.2 Composition

La structure d'une ontologie est composée, in fine, d'un ensemble de concepts associé à un ensemble de propriétés caractérisant ces concepts. Cette structure, représentant le squelette, de l'ontologie est organisée dans la **T-Box** (T pour terminologique). Les assertions d'individus et de rôles dans l'ontologie sont regroupées dans la **A-Box** (A pour assertion). On parle de *peuplement* de l'ontologie, bien

qu'il soit tout à fait commun, selon les applications, de travailler sur des ontologies non-peuplées.

Les concepts Les concepts sont organisés en une hiérarchie, appelée *hiérarchie de concepts* ou encore *taxonomie*. Nos deux concepts *Personne* et *Lieu* se trouveraient au même niveau de cette hiérarchie car aucun n'est plus général que l'autre. Ils expriment juste des notions différentes. En revanche, si l'on introduit par exemple un troisième concept, *Politicien*, il aurait sa place sous le concept *Personne*, tout politicien étant également une personne. Cette hiérarchie est régie par une relation de *subsomption*. On dit qu'un concept A subsume un concept B lorsque A est plus général que B, c'est à dire que tous les éléments de B sont également des éléments de A. Tous les concepts d'une ontologie sont subsumés par un concept abstrait plus général, auquel il est souvent fait référence sous le terme *Thing*, comprenant tout ce qui est. Les concepts représentent des ensembles d'individus (ou entités) aux caractéristiques communes. Nous pouvons par exemple dire que *Barack Obama* est une instance de *Personne* et *Hawai* est une instance de *Lieu*.

Les relations Les relations d'une ontologie s'expriment selon la même pratique que celle introduite par RDF, c'est à dire à base de triplets *(sujet, propriété, objet)*. Elles peuvent être de deux types. Ce sont soit des relations de subsomption, auquel cas elles marquent le caractère hiérarchique de la relation entre deux concepts (souvent représenté par le terme *is-a*), soit des relations sémantiques. Les relations sémantiques permettent de spécifier la nature d'une connexion entre deux concepts, ou plutôt entre deux individus appartenant à ces concepts. Dans notre exemple, la relation *né-à* connecte le concept *Personne* au concept *Lieu* par un lien spécifiant explicitement le fait d'*être né quelque part* (exemple : *(Barack Obama, né-à, Hawai)*).

Les attributs Les concepts peuvent également être enrichis d'attributs. Contrairement aux relations sémantiques connectant les concepts les uns aux autres, les attributs permettent d'associer une valeur littérale à un individu. Ils s'expriment cependant d'une manière similaire, sous la forme d'un triplet dans lequel l'objet est substitué par un littéral. La propriété *né-en* permettrait par exemple d'attacher au concept *Personne* une date de naissance, exprimée dans un format donné.

Les concepts, les relations et les attributs peuvent être contraints de plusieurs manières. Il est par exemple possible d'indiquer explicitement que deux concepts sont distincts l'un de l'autre, c'est à dire qu'aucun individu ne pourra être à la fois instance des deux. Il est aussi possible de contraindre la nature du sujet et de l'objet d'une propriété en utilisant les opérateurs *domain* et *range* introduits par RDFS. Les propriétés *né-à* et *né-en* auraient toutes deux comme domain le concept *Personne*, tandis que la première aurait pour range *Lieu* et la seconde un type de données, tel que *string* ou *date*. Les relations peuvent également avoir des propriétés telles que la

symétrie ($\{A$, *marié-à*, $B\}$ implique $\{B$, *marié-à*, $A\}$), la transitivité ($\{A$, *ancêtre-de*, $B\}$ et $\{B$, *ancêtre-de*, $C\}$ implique $\{A$, *ancêtre-de*, $C\}$) ou encore l'unicité. On parle pour cette dernière de relation *fonctionnelle*. Elle symbolise le fait que l'objet d'une telle propriété est unique. Un individu ne peut, par exemple, n'être né qu'à un seul endroit. Si plusieurs relations *né-à* lient une instance de personne à plusieurs instances de lieu, un moteur d'inférence en déduira que ces dernières représentent en fait le même lieu. A noter que cette propriété peut être inversée (on parle alors de relation *inversement fonctionnelle*), le sujet de la relation faisant alors l'objet de l'unicité.

De tels raisonnements sont possibles car, contrairement à RDFS, le langage OWL est basé sur la famille des logiques de description. L'ensemble des contraintes qu'il est possible d'utiliser dans la spécification d'une ontologie est dépendant du langage choisi pour sa construction, et notamment de l'expressivité de la logique qu'il implémente. Cette dernière sert en effet de cadre aux raisonnements conduits sur une ontologie et conditionne donc la quantité d'information et de connaissance qu'il sera possible d'inférer. Plus la logique utilisée est expressive et plus il est possible de spécifier des conceptualisations complexes, le prix à payer étant une augmentation de la complexité et donc du temps nécessaire au raisonnement.

La première version de OWL s'est ainsi vue déclinée en trois sous-langages à l'expressivité croissante : OWL Lite, OWL DL et OWL Full [McGuinness 2004]. Sa nouvelle version, OWL-2 [W3C 2012b], a récemment été introduite pour combler certaines lacunes d'expressivité liées à la conception de OWL-1 [Cuenca Grau 2008]. Le principe d'adaptabilité du langage au travail de modélisation a été conservé et enrichi grâce à l'introduction, en complément de OWL 2 DL et OWL 2 Full, des profils OWL 2 EL, QL et RL. Ces derniers sont destinés à optimiser les performances de l'ontologie en fonction du volume de concepts ou d'individus intégrés au modèle [W3C 2012c].

Les bases logiques de OWL permettent l'utilisation de mécanismes de requêtage et d'inférence plus complets que RDFS – bien que SPARQL reste complètemet compatible. Il est notamment possible de définir des concepts par des formules logiques (par exemple : *un père est un homme qui a des enfants*), laissant à la charge d'un moteur d'inférence l'instanciation de ces concepts par les individus correspondant à sa description. Il est également possible d'inférer des relations entre individus répondant à une description logique via l'application de *règles* (par exemple : *si Bob est le frère de Carole et que Carole est la mère de David, alors Bob est l'oncle de David*). De telles règles peuvent s'exprimer grâce à SWRL (pour *Semantic Web Rule Language*), inspiré de Datalog [Huang 2011], un langage de description de connaissances et de déduction basé sur la logique du premier ordre [Hitzler 2009], et soumis au W3C dans le but de devenir une extension officielle de OWL [Horrocks 2004].

On le voit, les fonctionnalités apportées par les ontologies reposent en grande partie sur la logique sous-jacente, dépendante du langage utilisé. La sous-section suivante introduit les bases des logiques de description ainsi que la manière dont

elles peuvent être utilisées pour la création d'ontologies.

2.4 Les Logiques de Description (LD)

2.4.1 Introduction

Les logiques de description forment une famille de langages adaptée à la description formelle de la connaissance d'un domaine, dans le but de la rendre manipulable par un système informatique. Le dictionnaire Larousse définit la connaissance comme étant "(l')action, (le) fait de comprendre, de connaître les propriétés, les caractéristiques, les traits spécifiques de quelque chose". Décomposons cette définition afin de poser clairement ce que nous entendons par connaissance. Commençons par la fin, le terme *quelque chose* renvoie aux entités du domaine que l'on souhaite décrire, initialement dépourvues de toute sémantique formelle. Cela pourrait par exemple être la notion de *mère* dans le cas où l'on souhaite modéliser le domaine des relations familiales. Ces notions primaires du domaine sont formalisées en concepts partageant les mêmes *caractéristiques*, les mêmes *traits spécifiques*, et reliés les uns aux autres par des *propriétés*. Une manière de formaliser le concept de *mère* serait alors de le définir équivalent au concept de *femme ayant au moins un enfant*. L'intelligence d'un système consiste en sa capacité à déduire des conséquences implicites à partir de connaissances définies explicitement. Si l'on formalise notre dernier exemple avec une logique de description, le fait de savoir que Alice est une femme et que Camille est son fils, associé à la connaissance de la définition du concept de *mère*, suffit à un système informatique implémentant cette logique pour déduire que Alice est une mère.

Les premiers formalismes à même de représenter des connaissances apparurent dans les années 1960 et 1970 avec les réseaux sémantiques [Quillian 1966] et les frames [Minsky 1974]. Ces deux paradigmes ont en commun de modéliser la connaissance sous la forme d'un réseau. Les ensembles d'individus correspondant au domaine modélisé sont symbolisés par les noeuds du graphe, tandis que les arcs reliant ces noeuds représentent les propriétés entretenues entres les types d'individus. Bien que réputés pour leur lisibilité, ils sont critiqués pour leur manque de précision dans la manière d'exprimer la connaissance [Johnson-Laird 1984], menant à des comportements différents et à un manque d'interopérabilité entre les systèmes. Les logiques de description ont été introduites pour remédier à ces problèmes. Les premiers travaux considérés comme fondateurs des logiques de description sont ceux de [Brachman 1985] qui introduisent le langage KL-ONE.

2.4.2 Différentes logiques fondatrices

Nous présentons brièvement dans cette sous-section les bases des logiques d'ordre 0 et d'ordre 1. Pas particulièrement parce qu'elles forment le socle sur lequel ont pu se construire les logiques de description mais parce qu'elles introduisent un certain nombre de symboles réutilisés par ces dernières.

2.4.2.1 Logique d'ordre 0

La logique d'ordre 0, ou logique booléenne, ou bien encore calcul des propositions, constitue la base du raisonnement logique mathématique. Elle permet de statuer sur la véracité d'une formule propositionnelle composée d'un ensemble de variables booléennes (x, y, etc.), ne pouvant donc prendre comme valeur que *vrai* ou *faux*, liées par cinq connecteurs grammaticaux [Richer 2008] :

- la négation $\neg$: si A est vrai alors $\neg A$ est faux ;
- la conjonction $\wedge$: $A \wedge B$, A et B sont vrais ;
- la disjonction $\vee$: $A \vee B$, A est vrai ou B est vrai ;
- l'implication $\rightarrow$: $A \rightarrow B$, si A est vrai alors B est vrai ;
- l'équivalence $\leftrightarrow$: $A \leftrightarrow B$, si A est vrai alors B est vrai, si B est vrai alors A est vrai.

où A et B sont des variables ou des formules propositionnelles.

Les constantes $\top$ et $\bot$ sont parfois utilisées comme symboles respectifs de *vrai* et *faux*. Les règles de priorité dans la lecture d'une formule sont classiquement assurées par les parenthèses.

On nomme *clause* une formule propositionnelle particulière composée uniquement de disjonctions de variables (ou de leur négation). $A = x \vee \neg y \vee z$ est une clause. On dit qu'une formule est sous forme normale conjonctive lorsqu'elle est composée d'une conjonction de clauses (exemple : $B = (x \vee y) \wedge (\neg x \vee \neg z)$). Toute formule peut être transformée en forme normale conjonctive selon les lois d'équivalence entre disjonction et conjonction introduites par De Morgan.

Le problème consistant à décider si une formule propositionnelle sous forme normale conjonctive possède une solution est connu sous le nom de problème SAT (pour boolean SATisfiability problem). Il a été démontré par [Cook 1971] que, quand le nombre de variables de la formule est supérieur ou égal à trois, SAT est un problème NP-complet. C'est à dire qu'il est possible d'en vérifier une solution donnée en un temps polynomial mais qu'il n'existe en revanche aucun algorithme à même d'en déterminer une dans un temps qui ne soit pas exponentiel. De plus, tous les problèmes NP peuvent être rapportés à SAT par une réduction polynomiale. C'est un problème majeur de la théorie de la complexité trouvant des applications dans beaucoup de domaines de l'informatique. La recherche d'algorithmes permettant de trouver une solution pour une formule ayant un nombre de variables raisonnable en un temps de calcul acceptable est donc logiquement l'objet de nombreux travaux [Zhang 2002, Prasad 2005].

Plusieurs méthodes permettent de prouver la validité d'une proposition. L'une des méthodes s'approchant le plus d'une déduction naturelle est la preuve à la Hilbert. Celle-ci s'appuie premièrement sur un ensemble de onze axiomes de la logique booléenne (voir tableau 2.1). Ces axiomes ont la particularité d'être des tautologies,

c'est à dire d'être vrais quelle que soit la valeur des formules qui les composent.

Nom	Axiome
Implication 1	$(X_1 \to (X_2 \to X_1))$
Implication 2	$((X_1 \to (X_2 \to X_3)) \to ((X_1 \to X_2) \to (X_1 \to X_3)))$
Négation 1	$(X_1 \to \neg\neg X_1)$
Négation 2	$(\neg\neg X_1 \to X_1)$
Négation 3	$((X_1 \to X_2) \to (\neg X_2 \to \neg X_1))$
Conjonction 1	$(X_1 \to (X_2 \to (X_1 \wedge X_2)))$
Conjonction 2	$((X_1 \wedge X_2) \to X_1)$
Conjonction 3	$((X_1 \wedge X_2) \to X_2)$
Disjonction 1	$(X_1 \to (X_1 \vee X_2))$
Disjonction 2	$(X_2 \to (X_1 \vee X_2))$
Disjonction 3	$(\neg X_1 \to ((X_1 \vee X_2) \to X_2))$

TABLEAU 2.1 – Onze axiomes tautologiques de la logique booléenne

Elle ajoute à ces axiomes le principe du *modus ponens* qui formalise le raisonnement mathématique naturel disant qu'à partir d'une formule A et d'une formule $A \to B$, il est possible de déduire B (voir équation 2.1).

$$\frac{A \quad (A \to B)}{B} \tag{2.1}$$

On dit qu'une formule propositionnelle A est prouvable à partir d'un ensemble de formules propositionnelles T si l'on a une suite finie $A_1, A_2, ..., A_n$ de formules propositionnelles telles que :

- A_n est égale à A
- Pour tout i
  - A_i est dans T, ou
  - A_i est un axiome de la logique booléenne, ou
  - A_i est obtenue par modus ponens de deux formules A_j et A_k avec $j, k < i$.

2.4.2.2 Logique d'ordre 1

La logique du premier ordre, aussi connue sous le nom de calcul des prédicats, permet la construction de formules plus riches que ce qu'il est possible d'obtenir avec le calcul propositionnel. Alors que ce dernier manipule essentiellement des variables booléennes indépendantes, la logique du premier ordre introduit la notion de prédicats dont la valeur de vérité dépend d'une ou plusieurs variables. La valeur du prédicat $Pair(x)$ dépend par exemple de la valeur de la variable x. Il sera vrai si x est paire, faux si x est impaire.

Au vocabulaire introduit par le calcul des propositions s'ajoutent :

- le quantificateur universel $\forall$: $\forall x$, pour toutes les valeurs possibles de x,
- le quantificateur existentiel $\exists$: $\exists x$, il existe une valeur pour x,
- un ensemble infini et dénombrable $\mathcal{V}$ de variables.

A noter que les quantificateurs ne sont applicables, en logique du premier ordre, que sur les variables. Les logiques d'ordre supérieur autorisent leur application sur les fonctions et les prédicats [Leivant 1994].

Au delà de ces nouveaux connecteurs, une logique du premier ordre est caractérisé par sa *signature* [Bournez 2013], notée $\mathcal{S} = (\mathcal{C}, \mathcal{F}, \mathcal{R})$ où :

- $\mathcal{C}$ est un ensemble de constantes. Exemple : 0, 42, π, etc.
- $\mathcal{F}$ est un ensemble de symboles de fonctions retournant des résultats en fonction d'une ou plusieurs variables passées en paramètre. Le nombre de variables d'entrée d'une fonction est donnée par son arité. Exemple : la fonction d'addition, $+$, d'arité 2 effectuant la somme de deux variables, la fonction racine, $\sqrt{()}$, d'arité 1 calculant la racine d'une variable, n'importe quelle fonction f d'arité n, etc.
- $\mathcal{P}$ est un ensemble de symboles de prédicats (ou relations) associant une valeur booléenne de vérité à une ou plusieurs variables, le nombre de variables étant, ici aussi, défini par l'arité de la relation. Exemple : $Pair$ d'arité 1, $=$ d'arité 2, etc.

Une formule du premier ordre s'écrit donc dans l'alphabet :

$$\mathcal{A}_\mathcal{S} = \mathcal{V}_\mathcal{S} \cup \mathcal{C}_\mathcal{S} \cup \mathcal{F}_\mathcal{S} \cup \mathcal{P}_\mathcal{S} \cup \{\neg, \wedge, \vee, \rightarrow, \leftrightarrow, \exists, \forall\} \tag{2.2}$$

Elle est syntaxiquement composée des formes suivantes.

Un terme d'une logique du premier ordre est un objet manipulé par cette logique [Rozière 2011]. L'ensemble $\mathcal{T}_\mathcal{S}$ des termes pour une logique de signature $\mathcal{S}$ est composé de :

- L'ensemble des constantes $\mathcal{C}_\mathcal{S}$,
- L'ensemble des variables $\mathcal{V}_\mathcal{S}$,
- L'ensemble des fonctions d'arité n de S pour lesquelles les arguments $t_1, ..., t_n$ appartiennent à $\mathcal{T}_\mathcal{S}$.

A noter qu'un terme ne contenant aucune variable est un terme dit *clôt*. Une constante est un terme clôt, $f(\pi)$ est un terme clôt, mais $x + y$ n'en est pas un.

Un atome, ou formule atomique, s'écrit sous la forme $P(t_1, ..., t_n)$ telle que P est un prédicat d'arité n appartenant à $\mathcal{P}_\mathcal{S}$ et $t_1, ..., t_n$ sont des termes appartenant à $\mathcal{T}_\mathcal{S}$.

Une formule se construit à l'aide de termes et d'atomes liés par des connecteurs et des quantificateurs. L'ensemble des formules de $\mathcal{S}$ se défini de manière inductive :

- Tout atome de $\mathcal{S}$ est une formule de $\mathcal{S}$,
- Si A est une formule, alors $\neg A$ est une formule,
- Si A et B sont des formules, alors $(A \wedge B)$, $(A \vee B)$, $(A \rightarrow B)$ sont des formules,
- Si A est une formule et x est une variable, alors $\forall x.A$ et $\exists x.A$ sont des formules.

Les variables apparaissant dans les formules peuvent être libres ou liées. On dit qu'une variable est liée lorsqu'elle est associée à un quantificateur, $\forall x$, $\exists x$. Elle n'a pas de valeur propre, elle est muette et peut être remplacée par n'importe quel autre symbole. Une variable libre est une variable non liée. Dans la formule $\exists x(x+y=z)$, x est une variable liée et y, z sont des variables libres.

Une formule qui ne contient que des variables liées est dite *close*. Exemple : $\forall y \exists x(x \leq y)$.

La méthode des tableaux sémantiques, utilisée pour le calcul de satisfiabilité d'énoncés en logique propositionnelle [D'Agostino 1999], peut également être utilisée en logique du premier ordre [Letz 1999]. Il s'agit d'un raisonnement par l'absurde permettant de déterminer qu'un énoncé est vrai en prouvant que sa négation est fausse. Par exemple pour prouver que $A \rightarrow (A \vee B)$, il faut prouver que $A \rightarrow \neg(A \vee B)$ est faux. L'ensemble de formules $\{A, A \vee B\}$ est donc placé sur la branche principale d'un arbre de vérité. Chaque formule est alors récursivement décomposée en formules plus simples selon les cas suivants :

- $A \wedge B$: A et B sont toutes les deux des conséquences logiques de la formule, elles sont donc ajoutées à la branche courante,
- $A \vee B$: soit A, soit B est une conséquence logique de la formule, deux nouvelles branches sont créées, une pour A et une pour B,
- $\exists x.A(x)$: il existe une valeur pour x satisfaisant $A(x)$. On utilise dans ce cas une *skolémisation*, x est instanciée par une constante c et la formule $A(c)$ est ajoutée à la branche courante.
- $\forall x.A(x)$: $A(x)$ doit être vraie pour toute valeur de x, x est donc remplacée par un terme t de la branche et $A(t)$ est ajoutée à la branche.

Lorsque apparaissent, sur une même branche, une formule et sa négation, la décomposition des formules de cette branche est stoppée et celle-ci est dite close. Le processus est effectué jusqu'à ce que plus aucune décomposition ne soit possible. Si toutes les branches sont fermées à ce moment là, la négation de la formule de départ est fausse, sa validité est donc prouvée.

Pour une introduction plus complète à la logique du premier ordre, le lecteur peut se référer aux articles de [Smullyan 1968, Fitting 1996].

2.4.3 Une famille de logiques

Le mécanisme de description utilisé par les logiques de description est hérité de la syntaxe de la logique du premier ordre. Les prédicats manipulés sont issus d'un vocabulaire spécifique au domaine modélisé. La description d'un concept, c'est à dire les conditions auxquelles doit répondre un individu pour être instance de ce concept, est effectuée par la combinaison des atomes du vocabulaire par différents opérateurs logiques. Dans la suite de cette section, nous adopterons la convention de notation dans laquelle A et B désignent des concepts atomiques, C et D des concepts décrits par des formules logiques et r un rôle atomique.

Le vocabulaire d'une logique de description est composé du triplet d'ensembles disjoints (N_C, N_R, N_I).

- Les éléments de N_C sont les *concepts* atomiques représentant des ensembles d'individus aux caractéristiques communes relatifs au domaine de connaissance modélisé. Ils sont comparables aux ensembles d'individus introduits notamment dans les réseaux sémantiques. Ils sont représentés par des prédicats unaires.
  Exemple : `Man(Bob)`, où `Bob` est un individu.
- N_R contient l'ensemble des rôles atomiques pouvant lier les éléments de N_C deux à deux. On rencontre aussi parfois le terme de *propriétés* pour les désigner. Un rôle est défini sous la forme d'un prédicat binaire prenant pour arguments les deux individus partageant cette propriété.
  Exemple : `hasWife(Bob, Alice)`.
- Enfin, N_I correspond aux individus de la base, aux données élémentaires venant s'inscrire dans la base de connaissances.
  Exemple : `Alice`, `Bob`.

Les logiques de descriptions possèdent plusieurs niveaux d'expressivité, et donc de complexité, en fonction des constructeurs qu'elles intègrent. Ces constructeurs sont les opérateurs logiques qui, associés au vocabulaire (N_C, N_R, N_I), permettent de définir de nouveaux concepts sous la forme $C \equiv D$, où C est le nom d'un concept et D une formule logique. Exemple : `Father(X)` $\equiv$ `Parent(X)` $\sqcap$ `Man(X)`. A noter que la conjonction ($\wedge$) et la disjonction ($\vee$) logiques se notent respectivement $\sqcap$ et $\sqcup$ dans la famille des LD. Les capacités de description d'une logique augmentent avec le nombre de constructeurs qu'elle intègre à son vocabulaire.

Une grande quantité de logiques différentes, on parle aussi parfois plus naturellement de langages, peuvent être générées par le regroupement de différents constructeurs. Ceux-ci sont représentés par des lettres qui, une fois concaténées, forment le nom de la logique qu'ils composent. Cette notation naturelle a été proposée par [Schmidt-Schauss 1991]. Elle a l'avantage de permettre d'associer un nom unique à chaque logique et, qui plus est, de rendre ce nom porteur d'information sur l'expressivité de la logique. La majorité des logiques concrètement utilisées sont dérivées du langage $\mathcal{AL}$ pour $\mathcal{A}$ttributive $\mathcal{L}$anguage, dont l'expressivité est considérée comme étant le minimum requis pour un quelconque intérêt pratique [Baader 2003]. Elle permet de décrire un concept arbitraire C en utilisant :

- Des concepts atomiques A
- Le concept universel $\top$
- Le concept absurde $\bot$
- La négation de concept atomique $\neg A$
- L'intersection de concepts $C \sqcap D$
- La restriction de valeur $\forall r.C$
- La restriction existentielle limitée $\exists r.\top$ ou $\exists r.\bot$

Le tableau 2.2 présente les différents constructeurs existants et leur signification.

Lettre	Signification
$\mathcal{E}$	Restriction existentielle complète Exemple : $\exists$`hasChild Daughter`
$\mathcal{F}$	Rôles fonctionnels Exemple : `hasBirthPlace`
$\mathcal{U}$	Disjonction de concepts
$\mathcal{C}$	Négation de concepts atomiques ou non
$\mathcal{H}$	Hiérarchie de rôles Exemple : `hasDaughter` est un sous-rôle de `hasChild`
$\mathcal{R}$	Conjonction de rôles
$\mathcal{R}^+$	Transitivité des rôles Exemple : `hasAncestor`
$\mathcal{O}$	Restriction sur les individus Exemple : `hasChild one-of {Jean, Julie, Antoine}`
$\mathcal{I}$	Inversion de rôle Exemple : `hasChild`$^{-1}$ signifie *est l'enfant de*
$\mathcal{N}$	Restriction cardinale Exemple : ≥ 2 `hasChild`
$\mathcal{Q}$	Restriction cardinale qualifiée Exemple : ≥ 2 `hasChild Female`

TABLEAU 2.2 – Constructeurs composant les logiques de description

La logique $\mathcal{ALC}$ (pour $\mathcal{A}$ttributive $\mathcal{L}$anguage with $\mathcal{C}$omplements), introduite par [Schmidt-Schauss 1991] est par exemple une extension de $\mathcal{AL}$, par l'ajout du constructeur de négation de concepts arbitraires $\mathcal{C}$. La logique $\mathcal{ALCN}$, reprend $\mathcal{ALC}$ en y ajoutant la notion de restriction cardinale. Il faut noter que tous les langages créés par l'intégration de nouveaux constructeurs ne sont pas nécessairement distincts deux à deux. Des concepts ayant la même sémantique peuvent en effet être exprimés de différentes manières et avec différents constructeurs. D'après les lois de De Morgan, la disjonction $\mathcal{U}$ et la restriction existentielle complète $\mathcal{E}$ sont notamment réductibles à la négation $\mathcal{C}$ associée respectivement à la conjonction et à la restriction de valeur présents dans $\mathcal{AL}$.

$$\exists r.C \equiv \neg\forall r.\neg C$$
$$C \sqcup D \equiv \neg(\neg C \sqcap \neg D)$$

Il est donc considéré qu'un langage qui permet d'exprimer la négation de concepts complexes intègre de fait les constructeurs de disjonction et de restriction existentielle complète, et réciproquement. Cela ce traduit généralement par la contraction du nom de la logique au dénominatif le plus court, $\mathcal{ALUE}$ et $\mathcal{ALUEN}$ devenant respectivement $\mathcal{ALC}$ et $\mathcal{ALCN}$. D'autres équivalences, telles que celle entre la restriction existentielle complète et la conjonction de rôles ou entre la négation de concepts atomiques et la restriction cardinale sur les rôles sont respectivement étudiées dans [Donini 1997] et [Baader 1996].

Des écarts à la notation peuvent être observés dans la dénomination de certaines logiques. C'est notamment le cas de $\mathcal{FL}$ et ses dérivées $\mathcal{FL}^-$ et $\mathcal{FL}_0$, se nommant ainsi car historiquement introduit par [Brachman 1984] comme *frame-based description language*, ou encore $\mathcal{S}$ qui sert d'abréviation à la logique $\mathcal{ALCR}^+$. A noter que OWL-Lite et OWL-DL correspondent respectivement aux logiques $\mathcal{SHIF}$ et $\mathcal{SHOIN}$ [Horrocks 2003], tandis que OWL 2 atteint l'expressivité de $\mathcal{SROIQ}$ [W3C 2012a]. L'expressivité des logiques parmi les plus courantes est donnée dans le tableau 2.3.

Constructeurs		**Logiques de description**			
Description	**Syntaxe**	$\mathcal{AL}$	$\mathcal{ALE}$	$\mathcal{ALCN}$	$\mathcal{SHOIN}$
concept le plus spécifique	$\top$	×	×	×	×
concept le plus général	$\bot$	×	×	×	×
négation d'un concept atomique	$\neg A$	×	×	×	×
négation d'un concept	$\neg C$			×	×
conjonction de concepts	$C \sqcap D$	×	×	×	×
disjonction de concepts	$C \sqcup D$			×	×
restriction existentielle limitée	$\exists r.\top$	×	×	×	×
restriction existentielle complète	$\exists r.C$		×	×	×
restriction de valeur	$\forall r.C$	×	×	×	×
restriction maximale	$\leq nr.C$			×	×
restriction minimale	$\geq nr.C$			×	×
hiérarchie de rôles	$r_1 \sqsubseteq r_2$				×
restriction sur les individus	$\{a_1, ..., a_2\}$				×
inversion de rôles	r^{-1}				×
transitivité de rôles	r^*				×

TABLEAU 2.3 – Constructeurs logiques pour les langages $\mathcal{AL}$, $\mathcal{ALE}$, $\mathcal{ALCN}$ et $\mathcal{SHOIN}$.

2.4.4 Sémantique

La sémantique d'une logique de description s'exprime en terme d'*interprétation*. Une interprétation est un couple $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ où $\Delta^{\mathcal{I}}$ est un ensemble non-vide, dénommé domaine de $\mathcal{I}$, et $\cdot^{\mathcal{I}}$ la fonction d'interprétation qui fait correspondre :

- chaque concept $C \in N_C$ à un ensemble $C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$,
- chaque rôle $r \in N_R$ à une relation binaire $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.

La sémantique des différents constructeurs est définie à partir de ces bases et synthétisée dans le tableau 2.4.

Constructeur	Sémantique
$\top$	$\Delta^{\mathcal{I}}$
$\bot$	$\emptyset$
$\neg C$	$\Delta^{\mathcal{I}} \backslash C^{\mathcal{I}}$
$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
$\exists r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y : (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
$\forall r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \forall y : (x,y) \in r^{\mathcal{I}} \rightarrow y \in C^{\mathcal{I}}\}$
$\leq nr.C$	$\{x \in \Delta^{\mathcal{I}} \mid \ \mid y : (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}} \mid \leq n\}$
$\geq nr.C$	$\{x \in \Delta^{\mathcal{I}} \mid \ \mid y : (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}} \mid \geq n\}$

TABLEAU 2.4 – Sémantique des constructeurs des logiques de description

La base de connaissance est notée $\mathbf{K} = (\mathbf{T}, \mathbf{A})$, avec $\mathbf{T}$ et $\mathbf{A}$ étant respectivement une T-Box et une A-Box définie. Un concept sera considéré comme cohérent, ou satisfiable, si et seulement si il existe une interprétation $\mathcal{I}$ pour laquelle $C^{\mathcal{I}} \neq \emptyset$. Dans le cas contraire, on dira que C est insatisfiable, c'est à dire que, quel que soit l'état du modèle, aucun individu de la base de connaissance ne sera instance de ce concept.

2.4.5 Raisonnement

La puissance des logiques de description ne réside pas essentiellement dans leur capacité à décrire de manière explicite la connaissance d'un domaine. Elles sont également capable d'effectuer des raisonnements afin de mettre en exergue les connaissances implicites du domaine ou de répondre à des requêtes. Ces raisonnements se divisent en quatre types d'inférence logique, étroitement liés les uns aux autres.

Satisfiabilité Comme évoqué précédemment, un concept C est satisfiable s'il existe une interprétation lui attribuant une extension non vide, insatisfiable dans le cas contraire. Si l'on définit par exemple un concept $C \equiv A \sqcap B$ et que A et B sont disjoints, alors C est insatisfiable car aucun individu ne peut à la fois faire partie de l'extension de A et de l'extension de B.

Subsomption La subsomption cherche à déterminer si un concept C est plus général qu'un concept D ou, en d'autres termes, si les propriétés de D sont également des propriétés de C. Si tel est le cas, il est dit que C subsume D, noté $D \sqsubseteq C$. En termes d'interprétation cela signifie que, pour tout modèle $\mathcal{I}$, tous les éléments de D sont également des éléments de C, $D^{\mathcal{I}} \subseteq C^{\mathcal{I}}$.

Consistance La consistance s'attache à vérifier la cohérence entre les assertions de la A-Box et la structure terminologique de la T-Box. Il s'agit de vérifier que **K** possède un modèle $\mathcal{I}$, tel que tous les concepts de **T** et toutes les assertions de **A** soient satisfaits dans $\mathcal{I}$. Si l'on a, par exemple, les concepts C et D notés disjoints dans **T** et que l'individu x est soit inséré, soit inféré de part ses propriétés comme instance de ces deux concepts dans **A**, on dit que **A** est inconsistante par rapport à $\mathcal{T}$.

Instanciation L'instanciation consiste à affecter chaque individu de la A-Box aux concepts auxquels il correspond en fonction des propriétés de l'individu et des caractéristiques des concepts. Une assertion de type `Man(Bob)` entrainera bien entendu directement l'instanciation de l'individu `Bob` au concept *Man*. Si l'on prend une deuxième assertion `hasChild(Bob,Carol)`, sachant que le domaine de la propriété *hasChild* est *Parent* et que $Father \equiv Parent \sqcap Man$, le concept *Father* sera également instancié par `Bob` sans que l'assertion `Father(Bob)` n'ait dû être effectuée.

Deux autres mécanismes d'inférence sont relevés par [Baader 2003], *l'équivalence* et la *disjonction* de concepts. La premier consiste à poser que deux concepts C et D sont équivalents si, pour toute interprétation $\mathcal{I}$, on a $C^{\mathcal{I}} = D^{\mathcal{I}}$, c'est à dire que tous les individus instances de C sont également instances de D. On note alors $C \equiv D$.

Le deuxième est l'opération intuitivement opposée, deux concepts étant considérés comme disjoints si, pour toute interprétation $\mathcal{I}$, on a $C^{\mathcal{I}} \cap D^{\mathcal{I}} = \emptyset$.

Les raisonnements effectués sur la **K** sont tous réductibles au problème de satisfiabilité.

- A subsume B si et seulement si $(B \sqcap \neg A)$ n'est pas satisfiable,
- a est une instance de A si et seulement si $\mathbf{K} \cup \{\neg A(a)\}$ n'est pas satisfiable,
- A et B sont équivalents si et seulement si $((A \sqcap \neg B) \sqcup (B \sqcap \neg A))$ n'est pas satisfiable,
- A et B sont disjoints si et seulement si $(A \sqcap B)$ n'est pas satisfiable.

Le méthode de preuve introduite par [Schmidt-Schauss 1991] pour la logique $\mathcal{ALC}$ est basée sur les tableaux sémantiques.

Hypothèse du monde ouvert Les ontologies fonctionnent traditionnellement avec l'hypothèse du monde ouvert (*Open World Assumption* ou *OWA*) en opposition à l'hypothèse du monde clos (*CWA*). Cela implique simplement que l'absence

d'information ne permet pas de déduire la négation de l'information. L'absence d'information est juste interprétée pour ce qu'elle est, une non-connaissance de cette information. Cela permet notamment aux ontologies d'être plus aptes à gérer l'incertitude et l'incomplétude des données pour la représentation d'un domaine et le raisonnement que, par exemple, les bases de données relationnelles.

Cette hypothèse peut toutefois mener à des déductions du moteur d'inférence pouvant sembler contre-intuitives. On pourrait par exemple vouloir définir un concept représentant les enfants issus d'une famille, au mieux, mono-parentale ainsi :

$$SingleParentedChild \equiv Child \sqcap \leq 1\ hasParent(Parent)$$

Un raisonneur n'affecterait pourtant aucun individu à *SingleParentedChild* car les instances de *Child* n'ayant qu'une seule relation *hasParent* ont peut-être d'autres parents dont on ne connait pas l'existence. L'hypothèse du monde fermé peut-être simulé localement via l'exécution d'une requête SPARQL comptant le nombre d'assertions dans la base de connaissance. Il est cependant plus sage de retourner le problème en utilisant des restrictions de cardinalité minimale. Les opérateurs d'égalité et de cardinalité maximale devant être utilisés comme des contraintes sur des concepts instanciés plutôt que comme des moyens de déduction.

2.5 Conclusion

Nous avons vu que la représentation des connaissances, et plus particulièrement les ontologies, sont des technologies très porteuses pour la formalisation, le partage et la réusabilité d'informations et de connaissances entre systèmes, notamment au travers de leur exploitation au sein du web sémantique. Ce n'est pour autant pas le seul cadre d'utilisation exploité par la communauté et plusieurs ouvrages [Staab 2009, Van Harmelen 2007] dressent une liste non exhaustive de cas où l'utilisation d'ontologies se révèle pertinente.

Il arrive que le besoin de formalisation des connaissances, de structuration et du partage des données soit local et restreint à un sous-ensemble d'organisations. Les hôpitaux génèrent par exemple un important volume de données hétérogènes. Celles-ci peuvent être produites à la fois par des machines et par des humains et revêtent généralement un caractère sensible. Plusieurs initiatives d'ontologies ont vu le jour afin d'organiser ce flux de données aussi bien au sein de l'hôpital [Kataria 2008], qu'entre établissements [Anand 2010]. Les ontologies sont également souvent exploitées pour leur hiérarchie de concepts en tant que taxonomies. Un domaine particulièrement actif à ce propos est celui de la bio-informatique [Stevens 2009], la complexité biologique des composants du vivant se prêtant particulièrement bien au jeu de la classification. The Gene Ontology [The Gene Ontology Consortium 2000] est un exemple de projet ayant rencontré un certain succès [Bada 2004] dans son initiative d'établir un vocabulaire partagé entre organismes.

De nombreux domaines ont pu bénéficier de l'apport des ontologies ou de techniques plus larges de représentation des connaissances, tels que l'ingénierie logicielle

[Gasevic 2009], la robotique [Chella 2002, Amigoni 2005, Levesque 2007], les systèmes multi-agents [Van der Hoek 2007] ou encore l'analyse et l'annotation d'images [Bannour 2011, Tousch 2012] dont un état de l'art est présenté en section 3.5.

Chapitre 3

Analyse de la bande dessinée

Sommaire

3.1 Introduction . . . 35
3.2 Des codes graphiques établis . . . 36
3.2.1 Un univers décrit en cases . . . 36
3.2.2 Restitution de la bande son . . . 41
3.3 Point de vue traitement d'images . . . 46
3.3.1 Extraction des cases . . . 46
3.3.2 Extraction des bulles . . . 47
3.3.3 Extraction du texte . . . 47
3.3.4 Localisation des personnages . . . 48
3.3.5 Vers une approche globale . . . 49
3.4 Point de vue représentation du domaine . . . 49
3.4.1 Des encodages pour le contenu . . . 49
3.4.2 A Comics Ontology . . . 51
3.5 Des approches hybrides . . . 52
3.5.1 Annotation consensuelle d'images . . . 53
3.5.2 Réduction du fossé sémantique . . . 55
3.5.3 Support pour l'analyse d'images . . . 60
3.6 Conclusion . . . 61

3.1 Introduction

Ce chapitre présente un état de l'art des travaux portant sur l'analyse de bandes dessinées dans une optique d'extraction et d'interprétation automatique du contenu. La première section propose une synthèse des éléments constitutifs d'une bande dessinée, de leurs codes graphiques et de leurs spécificités visuelles. Ces caractéristiques, inhérente au domaine de la BD, forment les fondations sur lesquelles les algorithmes de traitement d'images et les conceptualisations des connaissances peuvent être bâtis. La seconde section dresse ensuite un panorama des techniques de traitements d'image recensées dans l'état de l'art pour l'extraction des éléments graphiques à partir d'images numériques. La troisième section présente les différents formalismes permettant de représenter le contenu d'une bande dessinée ayant émergé ces dernières années. Enfin, une synthèse de l'apport des techniques de représentation des connaissances pour l'analyse d'images, de bandes dessinées mais également d'autres types, est proposée dans la dernière section.

3.2 Des codes graphiques établis

3.2.1 Un univers décrit en cases

Le but premier d'une bande dessinée est de raconter une histoire. Elle le fait en juxtaposant, dans un espace donné, des images, agrémentées ou non de texte. Mais la bande dessinée est aussi un média froid [McLuhan 1968], dans le sens où l'information fournie au lecteur est parcellaire et qu'une participation active de sa part est requise pour appréhender le message transmis dans sa globalité. En effet, si la bande dessinée est définie comme un art séquentiel [Eisner 1998], elle cultive également l'art de raconter des choses sans les dire ni les montrer, à travers l'*ellipse*. Scott McCloud, dans son ouvrage l'*Art Invisible* [McCloud 1993], décrit l'ellipse [1] comme le procédé mental instinctif du lecteur, permettant de compléter l'information incomplète fournie par une bande dessinée, notamment celle présente entre les cases.

Les cases, ou vignettes, sont les images, généralement délimitées par un cadre fermé, illustrant le récit mis en scène dans la bande dessinée. Elles représentent des instants figés d'une histoire continue. L'action se déroulant entre deux de ces instants doit être imaginée par le lecteur et ne peut qu'être suggérée par l'auteur. Le choix de ces instants et l'habileté avec laquelle ce qui n'est pas montré est suggéré n'est pas anodin et fera souvent la différence entre une œuvre de bonne ou de moins bonne qualité.

3.2.1.1 Le contenu des cases

Comme en cinéma et en photographie, le choix du cadrage est une problématique essentielle de la bande dessinée. Le plan et l'angle de vue sélectionnés sont cruciaux pour la mise en scène car ils définissent, par le placement de l'oeil-caméra du lecteur, la relation de celui-ci à la scène qu'il a sous les yeux. Chaque image est confectionnée dans le souci de susciter une réaction émotionnelle chez le lecteur.

Les plans en particulier servent à insuffler du rythme à la narration en variant les manières de présenter l'action. Chacun d'entre eux incite le lecteur à se concentrer de manière inconsciente sur un aspect particulier du récit en modulant l'aspect comique ou dramatique de la scène qu'il représente.

La littérature [Lainé 2007, Bergaglio 2004, McCloud 2006] en liste quelques uns, chacun servant un effet narratif particulier.

- **Le plan d'ensemble**, ou plan large, est un plan introductif, souvent utilisé comme transition entre deux scènes, traditionnellement en haut de page. Il a un rôle principalement descriptif, il est de ce fait généralement très détaillé, dessiné dans une case occupant une large partie de la planche, et contient peu de texte afin d'immerger le lecteur dans l'univers raconté. Il permet de localiser l'action dans l'espace et dans le temps afin que le lecteur puisse, sans que

1. Le terme employé par McCloud en version originale est *closure*. Le mot *ellipse* est une proposition du traducteur, à défaut de mieux.

ces informations lui soient rappelées dans la suite des cases, situer le contexte de l'histoire.

- **Le plan général** a un rôle comparable au plan d'ensemble. Plus resserré que ce dernier, il cadre et place dans le décor un groupe de personnages allant prendre part à l'action des cases suivantes. Il sert de transition et guide le lecteur de la case d'introduction vers le vif du sujet.

- **Le plan moyen**, ou plan américain, met en scène les personnages prenant part à la l'action d'une manière directe. Ils sont dessinés à la ceinture ou à mi-jambe, toutes les parties du corps servant de moyen d'expression (visage, bras...) étant visibles et détaillées. C'est le plan utilisé pour mettre en scène des dialogues, des confrontations, et pour faire évoluer la narration. Le lecteur est plongé au coeur de l'action, la « caméra » se situant à hauteur des protagonistes.

- **Le gros plan**, et très gros plan, servent à fixer l'attention du lecteur sur un élément en particulier. L'intention derrière l'utilisation d'un gros plan peut être purement narrative, celui-ci permettant d'illustrer une action qui serait imperceptible dans un plan plus large, ou bien à avertir le lecteur de l'importance de tel ou tel élément de la scène. Il peut également être utilisé pour renforcer le côté dramatique ou comique d'une scène en représentant avec intensité des émotions présentes sur un visage par exemple. L'élément représenté en gros plan occupe presque la totalité de la case, la case étant elle-même souvent de petite taille afin de renforcer l'effet de confinement lié à la proximité de la « caméra » au sujet représenté. Le gros plan peut être enrichi d'un effet de profondeur de champ donnant du relief à la scène en intégrant à l'arrière-plan un autre objet, un décor ou personnage.

L'alternance des angles de vue est un deuxième moyen pour briser efficacement la monotonie de illustrations. Chaque angle de vue a de plus une valeur évocatrice inconsciente chez le lecteur. Ils permettent d'exprimer un ensemble d'idées, de générer des impressions et des sentiments servant la narration.

- **En vue normale** l'oeil du lecteur se situe au niveau du sujet. L'effet produit est plutôt neutre, les conditions d'observation de la scène se voulant naturelles. Le choix du cadrage associé à cet angle de vue est alors important car c'est lui qui porte la majeure partie de l'effet psychologique du plan.

- **Une vue en plongée** désigne une scène vue depuis un point d'observation plus élevé que le sujet. Elle permet de rendre compte de l'immensité d'un lieu, de décrire de vastes décors, intérieurs ou extérieurs, et d'exprimer l'idée d'un déplacement et de la solitude. Ce type de point de vue donne également la sensation au lecteur de survoler la scène. Les personnages représentés sous cet

angle ayant tendance à être écrasés au sol, une relation de domination psychologique s'établit entre le lecteur et le sujet.

– **Une vue en contre-plongée** est l'exact opposé d'une vue en plongée, c'est à dire une scène vue depuis point d'observation plus bas que le sujet. Ce cadrage a pour effet de grandir les personnages représentés, leur procurant de ce fait une image de puissance et de supériorité sur leur environnement. L'oeil de la caméra étant tourné vers le ciel, c'est également l'occasion de mettre en valeur un personnage sur un fond uni et d'insister sur sa majesté à cet instant du récit.

– **L'alternance des cadrages en champ/contre-champ** n'est pas un angle de vue à proprement parler mais une façon d'associer deux angles de vue l'un à la suite de l'autre afin de produire un effet particulier. Cette association est notamment très efficace pour la mise en scène des dialogues, la caméra faisant successivement face à l'un ou l'autre des intervenants, ou toute autre situation opposant deux points de vue distincts. La règle cinématographique des 180 degrés [2] s'appliquant également à la bande dessinée [Mateu-Mestre 2010], l'effet du champ/contre-champ simule de fait une rotation du regard du lecteur. Le procédé peut donc également être employé pour mettre en scène des éléments inattendus, apparitions, surprises ou gags.

3.2.1.2 L'agencement des cases dans la planche

Si le contenu des cases est fondamentalement important et qu'un dessin et une mise en scène de qualité sont constitutifs d'une bonne histoire, la bande dessinée cultive aussi l'art de la suggestion. Les cases ne représentant que des arrêts sur image d'une histoire sensée se dérouler de manière continue dans la tête du lecteur, charge à l'auteur de faire en sorte que les parties imaginées de son histoire soient conformes à ses intentions. Il peut pour cela jouer sur la forme et sur le positionnement des cases au sein des planches composant son récit.

Les cases rectangulaires, presque carrées, et de taille modeste (environ un dixième de la planche) sont les plus répandues. Elles permettent une mise en scène assez neutre et laisse place au contenu pour faire progresser la narration. Le caractère générique qu'elles arborent du fait de leur utilisation intensive participe au fait que chaque écart fait à ce format entraîne implicitement une pause narrative ou marque un temps fort du récit sur lequel il convient de s'arrêter. L'étirement d'une case en largeur provoque par exemple une réaction chez le lecteur qui pourra, selon le contenu de la vignette, suggérer un allongement du temps qui passe (figure 3.1a) ou un sentiment de submersion (figure 3.1b).

La transformation contraire, c'est à dire un étirement vertical, génère en revanche une impression de profondeur [Duc 1983] (figure 3.2a) ou renforce la sensation de vertige (figure 3.2b), particulièrement lorsqu'elle est associée à une vue en plongée

2. Cette règle stipule que, dans une même scène, la caméra ne doit en aucun cas franchir la ligne imaginaire reliant les deux éléments principaux de la scène, sous peine de perdre le spectateur.

(a)

(b)

FIGURE 3.1 – (a) L'étirement de la case centrale accentue la sensation du temps qui passe. Crédits : Scott McCloud - L'art invisible - HarperPerennial. (b) La largeur de la case renforce l'impression d'encerclement. Crédits : Hugo Pratt - Capitaine Cormoran - Glénat.

ou contre-plongée.

Si la case rectangulaire reste un canon de la mise en scène de bandes dessinées, toutes les déformations se retrouvent dans la littérature et sont bonnes pour appuyer un propos. On peut souvent trouver, particulièrement dans les mangas, des cases au découpage oblique insistant sur le dynamisme d'une scène d'action, ou encore des cases dont la forme souligne le contenu (effet brisé, case ronde, etc.). Nous sommes tellement habitués à ce qu'une case soit proprement délimitée qu'une absence de cadre ou un ajustement de la case au bord de la page produit en soi un effet d'évasion et d'immersion particulier [McCloud 2006].

L'organisation des cases dans la planche revêt également son importance. Loin d'être anodine, elle se doit de composer avec les formats choisi pour la mise en scène afin de produire un flux de lecture fluide qui ne sortira pas le lecteur de son immersion dans l'histoire. Il existe trois types de composition permettant de guider le lecteur d'une case à l'autre.

- **La juxtaposition** de cases correspond à la forme élémentaire de la bande dessinée où les cases sont placées côte à côte dans l'ordre de lecture naturel. Elle compose le fameux « gaufrier » de douze ou neuf cases que l'on retrouve respectivement dans les bandes dessinées franco-belges et les comics américains. L'ordre de lecture correspond alors à la norme culturelle d'origine de l'œuvre, c'est à dire de gauche à droite puis de haut en bas en occident, ou de droite à gauche puis de haut en bas au Japon.

(a) (b)

FIGURE 3.2 – (a) La vue étroite et le personnage en arrière plan donne une impression de profondeur à la case. Crédits : Victor De La Fuente - Haggarth - Casterman. (b) L'étroitesse de la case renforce l'impression de hauteur introduite par la falaise. Crédits : Roger Lecureux et André Chéret - Rahan - Vaillant.

- **La superposition** désigne le cas où une case vient déborder sur une autre. Cela crée un effet de profondeur dans la planche, hiérarchisant les cases en fonction de leur puissance visuelle ou thématique. Bien que l'ordre classique de lecture s'applique de manière naturelle à ce type d'agencement, celui-ci peut être utilisé pour guider l'oeil du lecteur au sein d'une séquence volontairement déstructurée, la case débordant faisant alors office de flèche.

- **L'incrustation** fait référence à une superposition totale, c'est à dire lorsqu'une vignette est intégralement incluse dans une case plus grande. Cet effet est souvent utilisé pour mettre en avant un détail de l'action représentée dans la case principale, ou pour introduire une idée de simultanéité de l'action présentée dans chaque case.

Ces compositions sont des éléments dont nous nous servirons dans notre analyse de la mise en page des planches en section 7.2.

3.2.2 Restitution de la bande son

Alors que, comme nous l'avons vu, les images d'une bande dessinée ne sont pas de simples illustrations mais bel et bien des vecteurs narratifs très importants, elles sont toutefois, la plupart du temps, complétées par ce que l'on pourrait appeler la « bande son » de la bande dessinée.

Ce que nous qualifions de bande son correspond ici à l'ensemble du vocabulaire graphique permettant de rendre compte du son à travers l'image. Cela inclut notamment les dialogues entre les personnages, les pensées et la narration, mais également tout ce qui contribue à retranscrire l'univers sonore d'une scène.

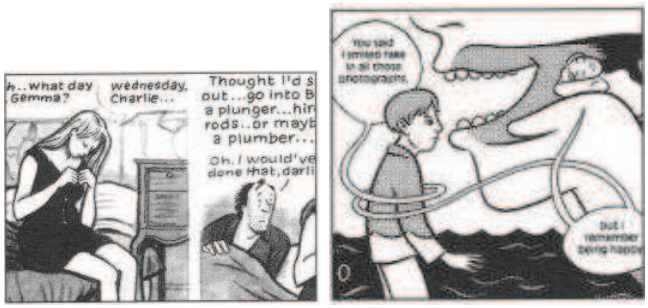
3.2.2.1 Les phylactères

Les *phylactères*, souvent appelés *bulles* et parfois *ballons*, sont les cadres, de différentes formes contenant le texte constitutif du récit. Ils ont une fonction purement narrative, laissant de côté tout ce qui a trait à la description de la scène. Leur usage s'est répandu dans la bande dessinée franco-belge dans les années 1920 à travers les œuvres de Saint-Ogan publiées dans le *Dimanche illustré* [CIBDI 2012]. Will Eisner, dans son ouvrage *La bande dessinée, un art séquentiel* [Eisner 1997], décrit les bulles comme une « tentative désespérée de capturer et de rendre visible un élément éthéré : le son ». Elles souffrent en effet du paradoxe d'avoir une apparence physique explicite dans le dessin, elles flottent au milieu de la scène, sans pour autant appartenir au même domaine de réalité. Elles prennent une place non négligeable de l'espace de dessin que les auteurs gèrent, une fois de plus, pour servir le récit. Certains auront tendance à minimiser leur impact visuel en supprimant partiellement ou totalement les contours (figure 3.3a), tandis que d'autres assumeront pleinement leur présence, les faisant même interagir avec le reste du dessin (figure 3.3b) [McCloud 2006].

Selon [Lainé 2010], les phylactères peuvent être classés en deux grandes classes, les bulles et les encadrés.

Les bulles sont les éléments graphiques, ressemblant à des ballons, servant à restituer les paroles des personnages. Elles peuvent à leur tour être réparties en trois catégories :

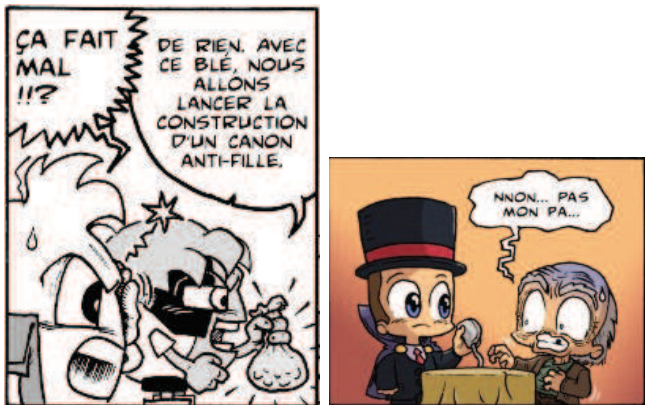
- Les bulles de parole servent à retranscrire la voix et les mots des personnages et sont celles que l'on retrouve le plus fréquemment. Elles sont généralement placées près de leur tête et présentent une excroissance en forme de flèche ou de pointe orientée vers la source du son. Lorsque cette flèche pointe vers l'extérieur de la case (on parle alors de bulle *hors-champ*), cela indique que la bulle est prononcée par un personnage présent dans la scène mais hors du cadre.



(a) (b)

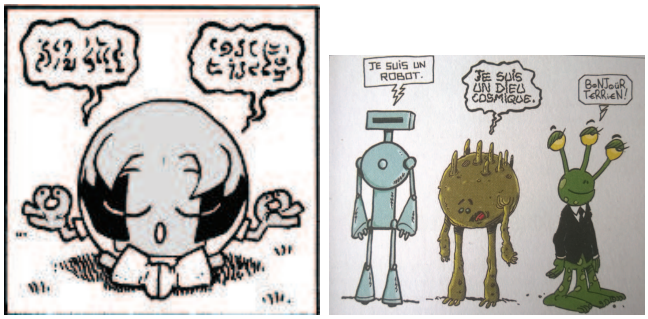
FIGURE 3.3 – (a) Le cadre des bulles s'efface pour réduire leur impact visuel sur le dessin. Crédits : Posy Simmonds - Gemma Bovery - Denoël. (b) Le caractère physique des bulles est assumé et elles interagissent avec le reste du dessin. Crédits : Hope Larson et Lucy Knisley - Letters from the bottom of the sea.

Le personnage en question apparait généralement dans les cases adjacentes. La forme des bulles n'est pas strictement contrainte et sert souvent à souligner l'état émotionnel ou l'origine du locuteur. Si la neutralité du dialogue est généralement représentée par une forme de bulle ovale, une bulle au contour « hérissé » (voir figure 3.4a) ou irrégulier (voir figure 3.4b) pourra, en fonction du contexte, être synonyme de douleur, de colère ou d'effroi. D'une manière générale, une bulle ayant une forme en rupture avec la convention implicite instaurée par l'auteur marque une différence, parfois dans le langage (figure 3.5a), parfois dans le type de locuteur (figure 3.5b).



(a) (b)

FIGURE 3.4 – Différents contours de bulles. (a) Contour hérissé marquant la colère ou la douleur. Crédits : Cyb - Bubblegôm Gôm Tome 1 - Studio Cyborga. (b) Contour irrégulier pour la faiblesse ou la peur. Crédits : Cyb - Le Magicien Loose - Studio Cyborga.



(a) (b)

FIGURE 3.5 – Un changement dans la forme des bulles peut parfois marquer (a) un changement de langage, ou (b) un changement d'espèce. Crédits : (a) Cyb - Bubblegôm Gôm Tome 1 - Studio Cyborga. (b) Illustration page 31 de [Lainé 2010].

– Les bulles de pensées matérialisent ce que pensent les personnages. Leur forme est conventionnellement « nuageuse » et leur queue est représentée par des cercles de plus en plus petits disposés en direction du penseur (voir figure 3.6).

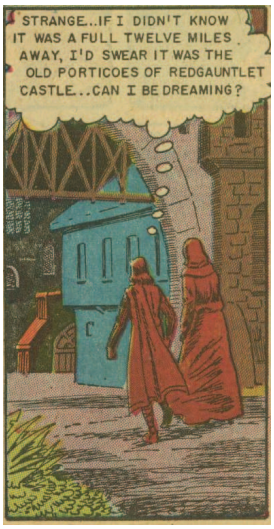


FIGURE 3.6 – Une bulle de pensée est caractérisée par sa forme nuageuse et par les petits cercles pointant le personnage. Crédits : Matt Fox - Chilling Tales 17 Geo - Youthful Magazines.

– Les hors-bulles sont des textes, généralement très courts, placés proche de la tête du personnage les récitant. Ils sont souvent utilisés pour matérialiser des interjections (« hein ? », « mais... », « gloup » etc.) ou des sons, des borborygmes (« grmblmblm », « gnnnnn » etc.).

Les encadrés se distinguent des bulles par leur forme rectangulaire, leur absence de queue et leur position dans les cases, généralement dans un coin et éloignée

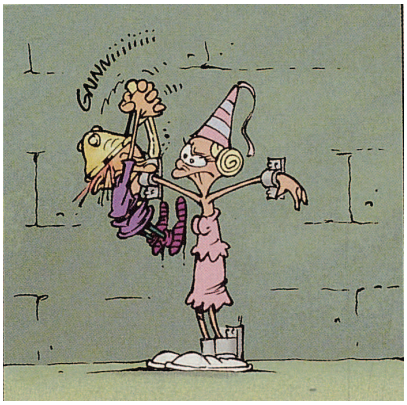
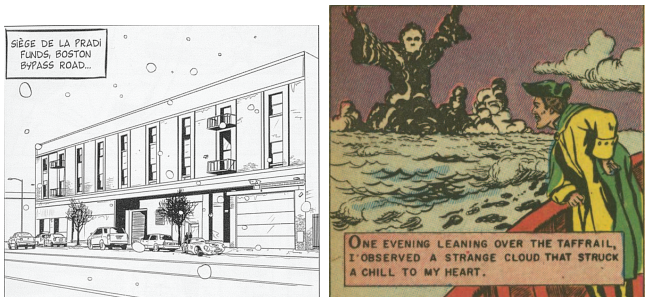


FIGURE 3.7 – Le hors bulle permet d'exprimer certains bruits ou certaines paroles prononcées de manière discrète. Crédits : Midam - Game Over, Walking Blork - Dupuis.

des personnages. [Lainé 2010] distingue deux sortes d'encadrés :

- Les récitatifs (voir figure 3.8a) sont de petits cadres contenant des textes relativement courts tels que *« Au même moment »* ou *« Un peu plus tard à Moulinsart »*. Ils retranscrivent la voix d'un narrateur omniscient ponctuant l'histoire de ses interventions pour aider le lecteur à localiser l'action temporellement et spatialement.

- Les voix off (voir figure 3.8b) diffèrent des récitatifs dans le sens où le texte encadré correspond aux paroles d'un personnage et non d'un narrateur. Le texte peut également être plus long que celui d'un récitatif, le personnage racontant ses souvenirs, un témoignage, etc.



(a) (b)

FIGURE 3.8 – Illustration des encadrés utilisés comme : (a) Récitatifs pour situer le lieu ou le moment de l'action. Crédits : Olivier Jolivet et Pascal Boisgibault - Boston Police : L'affaire Pradi - Clair de Lune. (b) Voix off déroulant la narration. Crédits : E. Allgor - Chilling Tales : Doomsday Ship - Youthful Magazines.

3.2.2.2 Le texte et les onomatopées

Le texte est le principal vecteur de transcription de la bande son d'une bande dessinée. Il matérialise à la fois les paroles, les pensées, la narration et les différents bruits que nous entendrions si en lieu et place d'une image nous étions devant un film. Si la forme des bulles donne au lecteur une indication sur le ton employé, ou l'atmosphère entourant un discours, les déformations graphiques du texte permettent également d'exprimer certaines variations dans la locution.

Pour des questions d'esthétisme et de lisibilité, le texte est souvent écrit en lettres capitales. L'élévation de la voix d'un personnage est illustrée par l'épaisseur du tracé des caractères, plus celui-ci est épais, plus la voix est forte [Lainé 2010]. De même, la taille du corps de caractère porte une indication sur le volume d'élocution des personnages, du chuchotement (petites lettres) au cri (grandes lettres).

La police de caractère utilisée a tendance à rester la même tout au long d'un album, dans un souci d'esthétique et d'uniformité. Des changements de police sont toutefois utilisés quelque fois pour marquer la spécificité de langage de tel ou tel personnage. C'est notamment visible pour le Grec et le Goth de l'album *Astérix Légionnaire* où les paroles de ce dernier sont par exemple transcrites en lettres gothiques (voir figure 3.9).

FIGURE 3.9 – Les différentes polices de caractères illustrent les différentes langues parlées dans la scène. Crédits : René Goscinny et Albert Uderzo - Astérix Légionnaire - Dargaud.

Les onomatopées sont les mots, directement incrustés dans l'image, dont la prononciation rappelle directement le son produit par un objet, une action ou bien un son naturel. Elles sont traditionnellement placées directement sur le dessin, au plus proche de la source du son qu'elles représentent [Bergaglio 2004]. La littérature s'est construit petit à petit un dictionnaire consensuel des onomatopées dont une liste est dressée par [Duc 1982]. Les bruits sourds, graves et étouffés sont évoqués par l'utilisation des voyelles « O » ou « U », tandis que les voyelles « I » et « E » ont plutôt tendance à rendre un son aigu [Dousset 2004] (voir figure 3.10).

Comme pour le texte des bulles, les déformations appliquées aux onomatopées permettent d'en moduler l'intonation ou l'intensité. Une variation de la taille des lettres fait immédiatement ressentir une variation dans le volume du bruit, alors

que le timbre ou la qualité du son est représenté par la régularité du dessin. Un son clair et propre est matérialisé par une onomatopées aux lettres lisses et bien définies, tandis qu'un son plus brouillé est par exemple transcrit par des lettres aux contours irréguliers ou ondulants [McCloud 2006].

(a) (b) (c)

FIGURE 3.10 – Plusieurs exemples d'onomatopées : (a) Un crissement de pneu, strident, proche de la source émettrice du son. Crédits : Olivier Jolivet et Pascal Boisgibault - Boston Police : L'affaire Pradi - Clair de Lune. (b) Un son de machine, transcrit dans une police adaptée, répétitif mais s'atténuant petit à petit. Crédits : G. Lubbin - Les bulles du labo - Doc en Stock (c) Un bruit bref, représenté par un mot prononcé rapidement, évoquant celui d'un liquide visqueux s'écrasant sur un solide. Crédits : Cyb - Les Aventures de MouetteMan et TeckelBoy - Studio Cyborga.

3.3 Point de vue traitement d'images

Comme nous venons de le voir, les images de bandes dessinées sont des documents complexes mêlant différents types d'objets afin de former un tout cohérent. Pour appréhender le contenu d'une page, il est nécessaire d'être en mesure d'en extraire le contenu de manière automatique. Dans cette section est dressé un panorama très rapide des différentes propositions émergeant dans la littérature. Cet aspect de l'analyse des images étant le cœur de la thèse de Christophe Rigaud, le lecteur intéressé peut se référer à [Rigaud 2014a] pour un état de l'art plus détaillé sur le sujet.

3.3.1 Extraction des cases

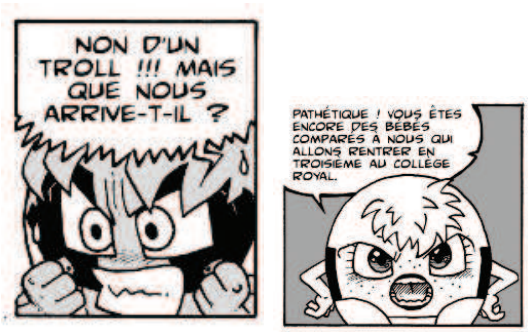
Les cases ont pour particularité d'être souvent rectangulaires et séparées les unes des autres par un espace vide. Ces caractéristiques sont exploitées par plusieurs méthodes s'appuyant sur les lignes de l'image [Li 2012b] ou sur la détection des bandes blanches [Duda 1972, Chung Ho 2007]. Ces méthodes ont pour faiblesse de ne pas fonctionner de manière optimale pour les cases n'ayant pas de contours.

Les approches basées sur les composantes connexes[3] ont partiellement résolu ces problèmes, au prix d'une baisse d'efficacité lorsque un objet graphique est à cheval sur plusieurs cases. Les travaux de [Arai 2010, Ho 2011] mettent notamment à profit le fait que les cases soient de gros objets visuels, occupant une large partie de l'espace de l'image et contrastant naturellement avec la couleur de fond de la page. La présence de contours nets, matérialisés par des traits noirs, est une particularité des images de bande dessinée. Elle est mise à profit lors de l'étape de binarisation. Une fois l'image binarisée, les cases correspondent naturellement aux plus grandes composantes connexes présentes dans l'image, la classification pouvant par exemple être effectuée par un K-means [Rigaud 2012].

3.3.2 Extraction des bulles

L'extraction des bulles a été légèrement moins étudiée que celle des cases. Plusieurs travaux [Arai 2011, Ho 2012] se sont toutefois attaqués au problème de la détection des bulles fermées. Ils se basent sur la détection de composantes connexes formant des blobs de couleur claire dans l'image. Celles-ci sont ensuite filtrées en fonction de leur taille et de la présence ou non de lignes de texte en leur sein.

Une méthode, basée sur le modèle des contours actifs [Kass 1988], a récemment été proposée par [Rigaud 2013b] afin de détecter les bulles ouvertes (voir figure 3.11b pour un exemple de bulle ouverte). Les contours des bulles classiques ont également été analysés afin de détecter la position et l'orientation de la queue pointant vers un éventuel personnage associé [Rigaud 2013a].



(a) Bulle fermée (b) Bulle ouverte

FIGURE 3.11 – Différence entre bulle ouverte et bulle fermée. Crédits : Cyb - Bubblegôm Gôm Tome 1 - Studio Cyborga.

3.3.3 Extraction du texte

La détection du texte dans une bande dessinée est un problème complexe car il est, d'une part, extrêmement difficile de déterminer, sans connaissance a priori, son

3. Une composante connexe est un ensemble de pixels adjacents formant une forme continue.

emplacement potentiel et, d'autre part, le fond de l'image n'étant pas uni, le texte est noyé dans un océan de composants graphiques hétérogènes. Plusieurs travaux ont permis de détecter le texte via une approche *top-down* en circonscrivant sa recherche à l'intérieur des bulles préalablement détectées [Yamada 2004, Sundaresan 2012]. D'autres approches sans a priori proposent de détecter le texte dans la page via des méthodes portant sur la classification des composantes connexes avec un SVM [Su 2011] ou un classifieur bayésien [Li 2013]. La minutie de certains dessins produit alors un grand nombre de faux positifs, les composantes connexes d'une lettre et d'un brin d'herbe pouvant être très proches. Les travaux de [Rigaud 2013c] proposent une méthode résolvant ce problème via un filtrage en trois étapes. Premièrement, les auteurs partent du principe que, dans un souci de lisibilité, les lettres sont représentées sur un fond très clair. Le contraste de la boite englobant la composante connexe doit donc être élevé. Deuxièmement, les lettres, dans la majorité des cas, ne sont pas sensées se chevaucher. Les boites englobantes des composantes connexes doivent donc être disjointes les unes des autres. Dernièrement, les mots étant écrits de manière rectiligne, les composantes connexes des lettres auront une tendance sensible à l'alignement, contrairement à l'organisation nécessairement plus chaotique des faux positifs.

3.3.4 Localisation des personnages

Plusieurs travaux se sont penchés sur la détection des personnages et notamment des visages, qu'ils soient en noir et blanc [Sun 2013] ou en couleur [Takayama 2012]. Les personnages de bande dessinée ne présentent pas toujours des visages réguliers et sont souvent soumis à des déformations graphiques spécifiques au média. Ils sont toutefois identifiables par les couleurs de leurs cheveux, de leur peau ou de leurs vêtements (voir figure 3.12).

FIGURE 3.12 – Exemple de déformation de personnage. Crédits : Vicky Portail-Kernel et Cédric Kernel - Prunelle, la fille du cyclope - Ankama

Des méthodes à base de graphes ont été utilisées pour détecter des motifs de couleurs répétés à travers les pages afin de les identifier comme étant les personnages principaux [Ho 2013]. Cette approche de recherche de motifs fréquents à été reprise par [Rigaud 2014c] pour retrouver toutes les occurrences d'apparition d'un personnage à partir de la labellisation d'une instance par un utilisateur humain.

3.3.5 Vers une approche globale

Les travaux présentés ci-dessus ont des angles d'attaques indépendants les uns des autres. Cependant, l'un des principaux objectifs de l'analyse d'images est d'être capable de comprendre le contenu représenté dans une image de manière automatique et dans toute sa globalité [Lamiroy 2014a]. Cela implique d'être en mesure de détecter et d'identifier les éléments présents dans l'image ainsi que d'en appréhender la sémantique. Si ces différentes tâches ont souvent fait l'objet de travaux indépendants [Mao 2003], il est toutefois envisageable de penser qu'elles peuvent se mener de front et que chacune d'elles peut aider à améliorer les résultats des autres. L'introduction de mécanismes de représentation des connaissances, en parallèle des outils d'analyse d'images, est alors nécessaire afin de formaliser le contexte de travail, les objectifs à atteindre et de poser un cadre structurant les différentes extractions effectuées sur les images [Lamiroy 2014b]. Plusieurs travaux ont adopté cette philosophie pour, entre autres, l'analyse de scènes naturelles [Hay 2008, Yao 2012, Li 2012a] et l'annotation d'images médicales [Hu 2003].

3.4 Point de vue représentation du domaine

Si l'on souhaite analyser le contenu d'une bande dessinée en s'appuyant sur des approches d'analyse holistique, il peut être utile de posséder une conceptualisation de sa structure, mais aussi des des interactions entre les éléments que le composent.

3.4.1 Des encodages pour le contenu

A notre connaissance, assez peu de travaux ont porté sur la conceptualisation du langage de la bande dessinée. Outre la proposition de webschéma *Periodical Comics* [W3C 2012d] et la plateforme *Grand Comics Database* [Klein 1994] qui sont toutes deux axées sur l'organisation bibliographique des œuvres, seules quelques initiatives portant sur la description de leur contenu ont vu le jour, sans toutefois être systématiquement les objets de publications académiques.

En 2001 a été proposée une formalisation XML, baptisée *ComicsML*, permettant de décrire le contenu d'une planche [McIntosh 2011]. Elle a été développée dans l'espoir de devenir un cadre pour la publication des webcomics alors naissants. En effet, la publication d'un webcomic se fait souvent directement sur la page web de l'auteur et la fréquence de mise à jour peut être très hétérogène en fonction des séries. L'ambition de ComicsML était donc de devenir un formalisme qui, si adopté par un grand nombre d'auteurs, permettrait de faire émerger de nouveaux usages de lecture des webcomics, notamment au travers d'outils exploitant les spécificités du langage. ComicsML permet de décrire une série de webcomics publiée sur le web, depuis ses aspects bibliographiques les plus généraux jusqu'au type de bulles ou à la forme du texte utilisée dans une case. La syntaxe, à base de balises, du

langage XML implique une approche hiérarchique dans la description des éléments composant l'image.

Le noeud racine, `comic`, décrit la collection à laquelle appartient un webcomic. Ses noeuds fils, `strip`, sont les instances de la collection, les planches réellement publiées. Ces balises contiennent notamment des metadonnées bibliographiques sur l'auteur ou encore la date de publication. Le contenu des cases est décrit au sein des balises `panel`, éléments fils des balises `strip`. Il comprend le type des bulles (pensées, dialogues, etc.), la typographie du texte (gras, souligné, etc.) et les personnages présents.

Le support du langage semble cependant été abandonné après la version 0.3. Une démarche partageant un certain nombre de points communs a été observée par [Morozumi 2009], mais appliquée aux mangas.

Une deuxième initiative, *CBML*, a vu le jour quelques années plus tard sous l'impulsion de [Walsh 2012]. Egalement basée sur XML, celle-ci est assez semblable dans sa philosophie à la proposition de [McIntosh 2011], bien que n'étant pas réservée à l'annotation de webcomics. Elle a en outre la particularité de s'appuyer sur la *"Text Encoding Initiative P5 : Guidelines for Electronic Text Encoding and Interchange"* [Text Encoding Initiative Consortium 2014]. La *TEI* est un cadre, proposé par le consortium du même nom, pour l'encodage de documents textuels, et notamment de documents textuels illustrés, dans le but de faciliter la recherche d'informations dans de grandes bases d'ouvrages encodés. *CBML* étend donc le vocabulaire de la *TEI* pour intégrer les notions spécifiques à la bande dessinée (cases, bulles, etc.), tout en réutilisant, autant que possible, l'encodage déjà existant.

Un autre formalisme, nommé *ACBF* (pour Advanced Comic Book Format), a été proposé par [Pastierovic 2011] et permet d'encoder certains aspects d'une bande dessinée numérique. Lui aussi basé sur le langage XML, ses ambitions se portent principalement sur l'amélioration de l'expérience de lecture[4]. Il se pose comme une évolution des formats d'archives de bandes dessinées que sont CBR et CBZ grâce à l'ajout de métadonnées sur plusieurs éléments. La notion de succession des pages et des cases y est notamment abordée, tout comme l'indication du type de bulles employé ou encore la langue dans laquelle a été écrit le texte. L'encodage de ce dernier est indépendant des images afin de faciliter d'éventuelles traductions d'œuvres. A la différence de ces prédécesseurs, *ACBF* permet également l'encodage de la position des cases et des zones de textes dans l'image, ceci afin de permettre d'effectuer des actions à la souris lors de la lecture.

Les différentes propositions présentées trouvent leurs applications respectives dans l'indexation, la recherche, et la lecture de bandes dessinées. Elles permettent d'encoder le contenu d'une BD de manière explicite à des degrés de granularité différents, en fonction de la tâche à laquelle elles sont destinées. Elles semblent tou-

4. Un lecteur compatible est téléchargeable sur la page web du projet : launchpad.net/acbf

tefois pensées pour une annotation entièrement manuelle, leur base XML commune rendant difficile l'application de mécanismes de raisonnement à même d'inférer de nouvelles informations sur les éléments annotés.

3.4.2 A Comics Ontology

À la date du lancement du projet eBDtheque [5], il n'existait pas, à notre connaissance, de véritables ontologies permettant de représenter le domaine de la bande dessinée et pouvant servir de bases à nos travaux. Dans le courant de l'année 2012, une proposition de [Rissen 2012] a toutefois été publiée sur le web. Bien que postérieure à la création de notre propre conceptualisation (présentée dans le chapitre 4), elle possède certaines similitudes avec cette dernière.

L'auteur s'est concentré sur la conception d'une ontologie permettant de décrire le contenu d'une bande dessinée plutôt que ses informations bibliographiques. La hiérarchie de concepts intègre les principaux éléments constituant une planche (représenté par le concept de *Card*), à savoir les cases (*Panel*), les bulles (*Balloon*) et les personnages (*Character).*

Le concept de bulle est spécialisé en cinq concepts afin de représenter différents exemples d'utilisation des phylactères, à savoir les bulles de dialogue, de pensée, d'exclamation, de chuchotement et de diffusion (lorsque le son est émis par une télévision ou un poste de radio par exemple). La notion de question-réponse entre deux bulles est représentée par la propriété *in-reply-to*, de même que la liaison graphique entre deux bulles est modélisée par la propriété *linkedBalloon.*

Le texte de la planche n'est pas à proprement parler représenté par un concept dédié. Lorsque celui-ci est véhiculé par une bulle, il est représenté comme étant un attribut de cette bulle et s'exprime à travers une propriété *xxx-content* où *xxx* correspond au type de bulle concerné (par exemple *exclamation-content.* Les onomatopées et les encadrés narratifs sont représentés dans une autre branche de la hiérarchie, respectivement par les concepts *Effect Caption* et *Narrative Caption*, subsumés par le concept *Caption.* Le texte en lui même est, de la même manière que pour les bulles, représenté par un littéral, attribut des instances de ces concepts.

La notion de personnage est représentée par le concept *Character* importé de l'ontologie OntoMedia [Jewell 2005]. Cette dernière a pour ambition de proposer un cadre à la modélisation des interactions entre médias. Les personnages peuvent être reliés aux bulles qu'ils émettent ou qui leur sont destinées par les propriétés *source* et *directed-at.*

Les éléments sont liés les uns autres par des propriétés aux *domain* et *range* contraints. Les bulles, encadrés et onomatopées doivent se trouver dans une case, tandis qu'une case doit se trouver dans une page. L'ordre des cases, par rapport à la page, et l'ordre des bulles, par rapport à la case, sont respectivement représentés par les attributs *panelOrder* et *balloonOrder.*

5. l3i.univ-larochelle.fr/eBDtheque

Plusieurs éléments d'annotation sont également intégrés à la conceptualisation. Il est possible de spécifier le lieu dans lequel se déroule l'action à travers un concept importé du vocabulaire RDF géographique proposé par [Brickley 2006]. Des informations additionnelles peuvent être intégrées grâce aux concepts *Note* et *Link*, le premier permettant d'ajouter un texte descriptif à une case ou une bulle, le deuxième renvoyant vers une ressource du web sous forme d'URL.

Bien que développée dans une perspective n'étant pas celle de l'analyse automatique de planches, l'ontologie proposée par Rissen comporte des éléments adaptés à cette tâche, que nous retrouvons d'ailleurs dans nos choix de conceptualisation présentés au cours du chapitre 4.

La hiérarchie de concepts de l'ontologie ainsi que la description de ses propriétés sont détaillées en annexe A.3.2.

3.5 Des approches hybrides

L'utilisation de connaissances a priori pour épauler la tâche d'analyse d'images n'est pas une initiative récente. Un panorama en a été dressé par [Crevier 1997], regrettant alors la trop grande hétérogénéité des approches, la spécificité des applications et le manque de standardisation pénalisant la compatibilité et le caractère réutilisable des méthodes développées. Les avancées effectuées depuis sur la représentation et le partage de connaissances, notamment grâce à l'émergence des ontologies, remettent une partie de ces conclusions en question. De nombreux travaux ont été menés, mettant à profit les capacités de description et de raisonnement des ontologies pour l'analyse, l'interprétation, l'annotation et la recherche d'images. A notre connaissance, très peu d'entre eux ont cependant été appliqués au domaine de la bande dessinée pour l'instant. Cette section a pour objectif de passer en revue, sans prétention d'exhaustivité, les avancées réalisées ces dernières années en analyse d'images grâce aux techniques de représentation de connaissances. Nous illustrons ces approches à travers la présentation de différents travaux.

Il est possible de catégoriser ces travaux de plusieurs manières, par le type de technologies qu'ils mettent en œuvre, par la place qu'occupe la phase de représentation des connaissances et de raisonnement dans le processus d'analyse ou encore par l'objectif poursuivi par l'étude. Nous avons choisi de dégager trois cas de figure particuliers, inspirés d'un panorama effectué par [Hudelot 2008a]. Premièrement, le cas où la connaissance modélisée est au service d'une annotation manuelle des images ou de zones dans ces images. Deuxièmement, le cas où elle permet une annotation automatique ou semi-automatique de ces mêmes éléments. Pour ces deux cas, l'utilisation de la connaissance intervient généralement en aval de l'analyse visuelle de l'image. Enfin, le cas où la connaissance a priori est au service du traitement d'images, soit par une analyse en amont des opérations à effectuer, soit par un feedback fourni aux algorithmes sur la concordance des extractions avec la connaissance modélisée.

3.5.1 Annotation consensuelle d'images

Lorsque l'on recherche une image, par exemple sur internet, l'idéal est de pouvoir décrire avec des mots le contenu que nous souhaitons voir apparaitre dans cette image. Le moteur de recherche de Google a popularisé cette pratique au début des années 2000. Les résultats obtenus, plus ou moins pertinents, sont basés sur des annotations extraites de l'analyse du texte proche de l'image sur une page web, du nom du fichier ou encore des liens s'y rapportant. Le contenu des images n'est cependant pas formellement identifié et il n'est pas rare que les résultats retournés par le système soient pollués par l'hétérogénéité sémantique de certains termes comme l'illustre la figure 3.13. Comment savoir dans quel cas le mot « avocat » se rapporte au fruit ou à la profession. Une personne cherchant le mot « java » souhaite t-il des illustrations de la danse, de l'île ou du langage informatique ?

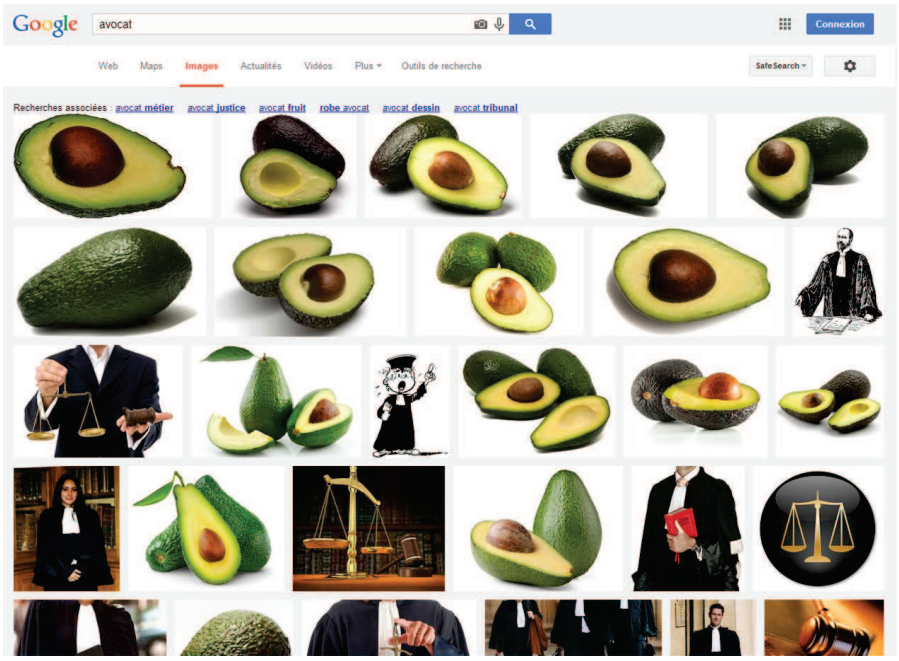


FIGURE 3.13 – Exemple d'hétérogénéité sémantique lors d'une requête par mot clé.

Les ontologies de description sont des vocabulaires structurés permettant de répondre à ce problème d'hétérogénéité. Parmi elles, les taxonomies regroupent des termes en fonction de leur signification à travers des relations de subsomption (« une voiture est un type de véhicule »). Les thésaurus étendent la portée descriptive des taxonomies en introduisant également la synonymie et les liens de proximité sémantique entre termes (appartenance au même champ lexical). L'utilisation du vocabulaire SKOS [Miles 2005] est recommandé par le W3C pour leur création. De tels outils permettent de mettre à disposition des vocabulaires communs, consensuels et désambiguïsés pour l'annotation d'images, pouvant éventuellement être combinés

les uns aux autres [Hollink 2003]. De plus, le caractère hiérarchique de l'organisation des termes d'un thésaurus permet de répondre à des requêtes nécessitant un raisonnement de subsomption logique. Une requête portant sur des images de véhicules pouvant alors avoir comme ensemble de réponses, en fonction de l'organisation sous-jacente du vocabulaire utilisé, des images contenant des instances d'hyponymes de véhicule, c'est à dire des concepts moins généraux (voitures, vélos, camions, etc. - cf. section 2.3.2).

Dès le début des années 2000, [Schreiber 2001] a proposé une méthodologie d'annotation de photographies, basée sur la combinaison de deux ontologies. La première, indépendante du sujet des photos à annoter et donc réutilisable dans d'autres contextes, modélise les concepts relatifs à la tâche d'annotation elle-même, comme par exemple la date et le lieu de prise de la photo, sa résolution, le nom du photographe ou encore le sujet de la photo. Ce dernier élément est le point d'ancrage qui permet de lier cette ontologie à la seconde, à laquelle échoit la tâche de représenter le domaine des photos – des photos de singes dans l'exemple donné. La portée du vocabulaire représenté s'étend des différentes races de singes à leur lieux d'habitations en passant par des termes décrivant la posture ou l'état des primates.

Le vocabulaire d'un thésaurus peut potentiellement devenir très volumineux, sa construction elle-même devenant alors une tâche fastidieuse. Afin de contourner ce problème, et d'assurer le caractère consensuel du thésaurus, certains travaux s'appuient sur WordNet [Hanbury 2008]. WordNet est une base de données lexicale, en ligne[6], regroupant une grande partie des noms, verbes, adjectifs et adverbes de la langue anglaise en ensembles de synonymes appelés synsets [Miller 1990]. Ces ensembles sont reliés les uns aux autres au sein d'un graphe sémantique où les arcs peuvent représenter des relations hyperonymiques (subsomption), méronymique (fait partie de) ou antonymique (contraire). Les travaux de [Zinger 2005] et de [Marszalek 2007] utilisent par exemple cette approche. Le premier extrait de WordNet un thésaurus de 24000 termes correspondant à des objets qu'il est possible de représenter visuellement (laissant de côté les concepts abstraits comme le bonheur, l'amertume, etc.), constituant ainsi une base d'annotations pour des images au contenu hétérogène. Le second extrait un thésaurus permettant d'annoter la base VOC [Everingham 2005], mettant en avant les bénéfices apportés en termes de temps de calcul et d'abstraction dans la classification des images de la base.

Les travaux de [Hyvonen 2002] illustrent, à travers la présentation d'une ontologie pour la description de photos du musée de l'université d'Helsinki, l'importance cruciale d'un vocabulaire formellement défini, non seulement pour la recherche d'images mais également pour la phase d'annotation elle-même. C'est particulièrement le cas lorsque les notions impliquées sont complexes ou lorsque la personne effectuant l'annotation n'est pas experte du domaine. Un vocabulaire structuré permet alors d'éviter des erreurs d'interprétation.

6. wordnet.princeton.edu

Toutefois, l'annotation d'images avec des termes structurés qualifiant le contenu représenté dans son ensemble n'est pas toujours satisfaisant. En effet, il peut être plus intéressant pour certaines application de localiser ses annotations dans l'image. C'est l'approche retenue notamment par [Hu 2003] dont l'étude porte sur l'analyse de radiographies pour le dépistage du cancer du sein. Des régions d'intérêts de l'image sont annotées selon leurs caractéristiques visuelles, telles que leurs dimensions, leur forme et leur texture. Les capacités de raisonnement des logiques de description sont ensuite mises à profit pour classifier les instances annotées en différentes pathologies.

Les travaux de [Maillot 2004] proposent une approche plus générique via une ontologie de concepts visuels (*Visual Concept Ontology*) indépendante d'un quelconque domaine d'application. Des experts d'un domaine (le pollen dans l'exemple donné par l'étude) sont invités à annoter des zones d'intérêt de l'image en utilisant des termes issus d'un vocabulaire accessible fourni par l'ontologie. Ces termes permettent de qualifier des caractéristiques visuelles, telles qu'une couleur ou une forme, auxquelles l'ontologie fait correspondre les descripteurs numériques permettant leur identification par un algorithme de traitement d'images.

Un état de l'art détaillé sur l'apport des vocabulaires structurés pour la résolution du problème de l'annotation sémantique d'images a été dressé en 2012 par [Tousch 2012]. Malgré l'impact bénéfique des ontologies, l'annotation manuelle d'images reste quoi qu'il en soit une tâche longue et fastidieuse. Le contenu visuel de l'image, l'information portée par ses pixels, leur couleur, leur agencement, n'est que rarement prise en compte. C'est pourtant sur celle-ci que se repose en grande partie l'analyse visuelle effectuée par un observateur humain. Il est donc naturel de penser pourvoir s'en servir pour reconnaitre le contenu d'une image et ainsi en automatiser l'annotation. Le fait est que l'organisation et la valeur des pixels seules ne sont pas suffisantes pour permettre une interprétation automatique. Cette distance présente entre les caractéristiques visuelles des pixels et la signification qu'elles ont pour un observateur humain est connue sous le nom de *fossé sémantique*.

3.5.2 Réduction du fossé sémantique

Le problème de réduction du fossé sémantique est souvent associé à la tâche d'analyse d'images mais on le retrouve dans tous les domaines de traitement automatique des données (voir exemple de l'analyse d'une phrase en 2.2.2). Il est défini par [Smeulders 2000] comme le manque de concordance entre l'information qu'il est possible d'extraire automatiquement de données (de données visuelles dans le cas qui nous préoccupe) et ce que ces mêmes données signifient pour un utilisateur dans un contexte particulier. Les composants atomiques d'une image étant ses pixels, des éléments ponctuels uniquement porteur d'une information de couleur, la recherche d'un lien avec la représentation concrète d'objets, et a fortiori de messages abstraits tels que des émotions, n'est pas trivial. De plus, l'interprétation du contenu d'une image n'est ni unique, ni absolue, et dépend des connaissances de la personne. Ainsi, chacun, en fonction de ses connaissances, pourra voir dans la photographie de la fi-

gure 3.14 une image en niveaux de gris, une bestiole, un insecte, un tardigrade ou un *hypsibius dujardini*.

FIGURE 3.14 – Différentes interprétations possibles pour une même image. Un truc ? Un animal ? Un *hypsibius dujardini* ?

Le domaine de la recherche d'images par le contenu (dénommé *CBIR* pour *Content-Based Image Retrieval*) [Gudivada 1995], c'est à dire la recherche d'images similaires à un exemple donné, s'est heurté de plein fouet au problème du fossé sémantique [Smeulders 2000], les caractéristiques pixellaires seules n'étant pas suffisantes pour juger de la proximité sémantique de deux images sans connaissances a priori. L'exemple présenté en figure 3.15 illustre les résultats d'une requête effectuée sur le moteur de CBIR de Google, à partir d'une image non référencée, de telle sorte qu'aucun mot clé ne lui soit préalablement associé. La plupart des images résultats contiennent bien des véhicules mais peut-être étais-je plus intéressé par des images de bandes dessinées illustrant une course poursuite avec la police.

La connaissance d'un certain contexte semble donc indispensable pour rendre possible l'interprétation d'images, comme l'a notamment souligné [Bannour 2011]. L'utilisation d'ontologies a été identifiée comme l'un des principaux moyens permettant de réduire le fossé sémantique [Liu 2007], aux côtés des méthodes d'apprentissage, de bouclage de pertinence, de l'utilisation des annotations du web et de l'utilisation des *semantic visual templates* [Chang 1998].

Plusieurs travaux se sont penchés sur la représentation des caractéristiques visuelles des images. Dans [Mezaris 2003], les auteurs proposent un système de recherche d'images par le contenu basé sur une ontologie intermédiaire (*Object Ontology*) permettant de qualifier l'aspect visuel de différentes zones de l'image par des termes simples. Les caractéristiques des zones (valeur des pixels, position, dimensions et forme) sont extraites et automatiquement associées à un ensemble prédéfini de valeurs telles que présentées sur la figure 3.16. Les objets à rechercher dans les images peuvent alors être décrits manuellement à partir des concepts de l'ontologie intermédiaire, cette dernière se chargeant de faire le lien avec les caractéristiques

FIGURE 3.15 – Exemple de fossé sémantique

visuelles des zones présentes dans les images. Un mécanisme itératif de bouclage avec l'utilisateur permet ensuite de retourner de nouvelles instances à partir des exemples annotés.

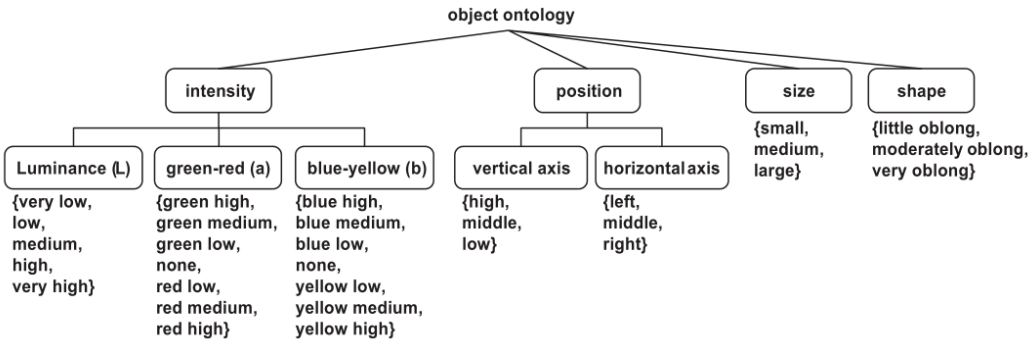


FIGURE 3.16 – Object Ontology de [Mezaris 2003]

Les travaux de [Jiang 2004] portent quant à eux sur la classification d'images d'œuvres d'art, d'une part en fonction de leur nature (photographies, peintures à l'huile, dessin traditionnel chinois, images générées par ordinateur), d'autre part sur les sentiments qu'elles dégagent (joie, colère, tristesse, etc.). Une ontologie fait correspondre les résultats de l'analyse des pixels (présence ou non de gradients nets sur les contours, saturation, luminance, etc.) à des concepts de haut niveau sémantique. Un utilisateur de ce système peut ensuite effectuer une requête par mot clé sur ces concepts afin de retrouver des images aux caractéristiques visuelles correspondantes et donc d'aspect similaire.

Les auteurs de [Petridis 2006] proposent une ontologie de descripteurs visuels (*Visual Descriptor Ontology*) permettant de lier un concept de haut niveau sémantique (un véhicule, un arbre, le ciel, etc.) avec un prototype de descripteurs visuels définis par la norme MPEG-7 [Sikora 2001] correspondant aux caractéristiques pixellaires de sa représentation dans une image ou une vidéo. Si l'application immédiate réside dans l'annotation de zones de l'image, l'objectif final est d'exploiter ces annotations pour répondre à des requêtes complexes, comme par exemple les instants d'une vidéo où une balle de tennis touche une ligne blanche. Ils mettent également en avant la généricité de leur approche quant au domaine d'application, la définition de nouveaux prototypes correspondant à de nouveaux concepts suffisant à passer d'un domaine à un autre.

L'aspect visuel d'un objet, que ce soit sa couleur, sa texture ou sa forme, n'est pas la seule information sur laquelle il est possible de s'appuyer pour l'identifier au sein d'une image. Les éléments reconnaissables d'une image sont souvent soumis à des contraintes de positionnement spatial imposés par la structure d'un document ou tout simplement par la réalité du monde physique. Si l'on recherche par exemple la signature de l'auteur d'une lettre numérisée, on aura plutôt tendance à l'assigner à une forme identifiée au bas d'une page plutôt qu'en haut. De même, l'annotation de deux grandes zones bleues aux caractéristiques visuelles similaires, issues de l'analyse d'une photo de plage, sera facilitée par la supposition préalable que le ciel tend généralement à occuper la partie supérieure de l'image. On peut voir un exemple d'une telle approche dans l'*Object Ontology* de [Mezaris 2003] illustré en figure 3.16, où le concept position fait le lien avec une position absolue de la zone dans l'image, ainsi que dans les travaux de [Zhang 2009]. L'intégration d'une connaissance spatiale a priori a été mis à profit dans plusieurs cas d'applications concrets tels que [Miron 2007, Hudelot 2008b, Coustaty 2011, Christodoulou 2012].

L'ontologie utilisée par [Coustaty 2011], dans le cadre de l'analyse d'images de lettrines, permet d'identifier, parmi un ensemble de zones segmentées dans l'image, celle correspondant à la lettre. Une lettre correspondant à la zone d'aire maximale, centrée dans l'image et contenant un nombre limité de trous dans leur forme (un trou pour la lettre A, deux pour la lettre B, etc.), un ensemble de règles modélisant ces contraintes sont intégrées dans l'ontologie afin de conduire la déduction.

Les travaux de [Hudelot 2008b] introduisent une ontologie (voir figure 3.17) permettant de représenter les relations spatiales topologiques (chevauchement des formes, inclusion de l'une dans l'autre, disjonction, etc.), directionnelles (à gauche de, en dessous de, etc.) et de distance (proche de, très loin de, etc.). La généricité des concepts de l'ontologie rend celle-ci indépendante d'un quelconque domaine d'application. La sémantique des concepts est exprimée sous la forme d'ensembles flous, la véracité d'un axiome ne s'exprimant plus alors de manière booléenne mais par une valeur de vérité comprise entre 0 et 1. L'introduction des principes de la logique floue aux méthodes de représentation des connaissances permet d'intégrer l'incertitude et l'imprécision inhérente à bien des applications du monde réel. Une proposition pour la représentation d'ontologies floues basée uniquement sur l'expressivité de OWL 2 est détaillée par [Bobillo 2011]. Cette sémantique floue est en revanche bel et bien dépendante du contexte car, comme le soulignent les auteurs, la notion *proche de* n'est par exemple pas la même pour un géographe analysant une carte (voir exemples de [Miron 2007]) que pour un clinicien examinant la radiographie d'un cerveau.

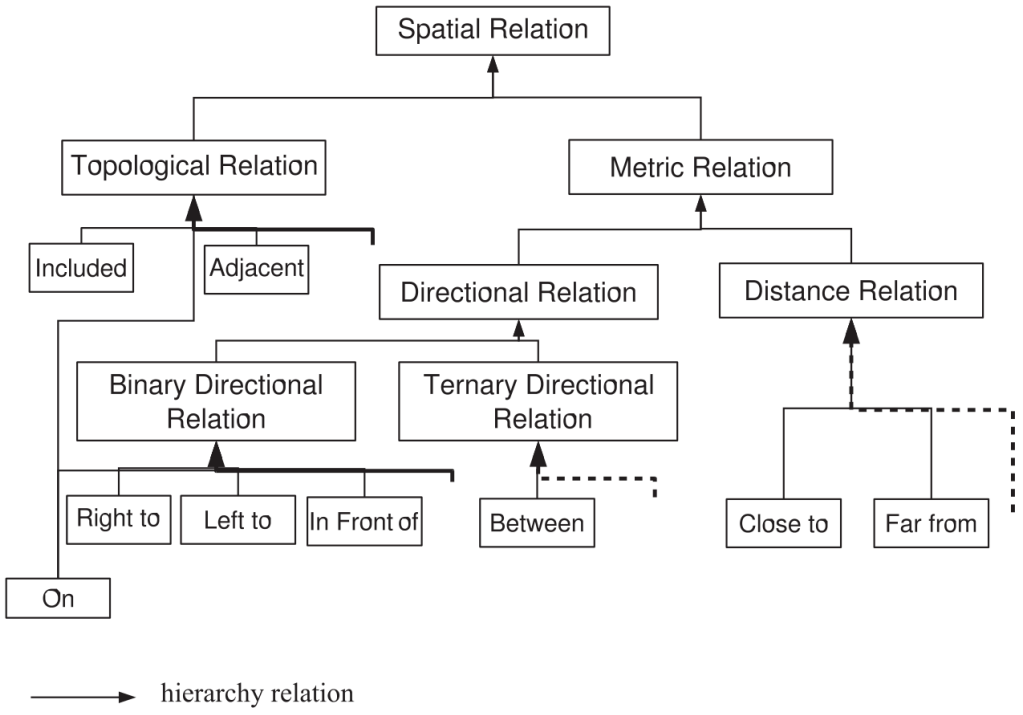


FIGURE 3.17 – Extraits de la hiérarchie de relations spatiales de [Hudelot 2008b]

Plus récemment, l'apport des ontologies pour l'annotation d'images satellitaires utilisant les techniques d'analyse d'images basée sur les objets (ou *Object Based Image Analysis* [Blaschke 2010]), notamment sur le côté consensuel et non ambigu des étiquettes, a été mis en avant par [Belgiu 2014].

Ces différents travaux témoignent à la fois d'un intérêt certain envers la représentation de données spatiales, ainsi que d'un besoin de standardisation. L'*Open Geospatial Consortium* a récemment accepté le formalisme GeoSPARQL [Battle 2012] comme un standard pour la représentation de données géospatiales et leur requêtage. Il permet de rendre compte de la forme des entités mais également de leurs relations spatiales topologiques. Les relations de direction et de distance ne sont pas modélisées directement mais restent potentiellement dérivables des coordonnées spécifiant la position des éléments. Les implémentations supportant les mécanismes de raisonnement introduits par GeoSPARQL au moment de l'écriture de ces lignes sont toutefois circonscrites aux technologies d'Oracle et aux entrepôts de données Parliament [7] et Strabon [8].

Le moteur d'inférence Pellet, développé par Clark&Parsia [9], a été enrichi de capacités de raisonnement spatial via son extension *Pellet Spatial* [Stocker 2009]. Ses capacités d'inférence se limitent toutefois aux relations topologiques de RCC-8 définies dans [Cohn 1997]. Il est en mesure de vérifier la cohérence d'une base et d'inférer de nouvelles relations à partir d'annotations topologiques déjà existantes mais il ne peut pas les déduire d'un ensemble de coordonnées. Récemment, [Christodoulou 2012] a proposé une extension de Pellet Spatial le rendant capable de raisonner également à partir de relations spatiales d'orientation, dénommées *Cone-Shaped Directionnal relations*, ou CSD. L'inférence de ces relations entre objets à partir de coordonnées reste cependant toujours hors de portée. L'optimisation de la composition des relations de RCC-8 a été étudiée dans [Le Ber 2002], mettant à profit les propriétés des treillis de Galois afin de réduire la combinatoire générée.

Le problème d'interprétation de l'action représentée au sein d'une image a été adressé par [Chen 2012]. L'étape de détection des zones de l'image contenant des objets d'intérêt est considérée comme triviale. La probabilité d'appartenance d'un objet à une classe donnée, parmi un ensemble contrôlé de possibilités, est estimée d'une part grâce aux caractéristiques visuelles intrinsèques de la zone correspondante, d'autre part grâce à la position relative des objets détectés les uns par rapport aux autres. Une ontologie (*Guide Ontology*) est ensuite utilisée pour instancier les liens entre les objets par des relations représentant des actions à travers l'optimisation d'une valeur d'énergie calculée à partir des contraintes de l'ontologie.

3.5.3 Support pour l'analyse d'images

Les systèmes d'interprétation d'images exposés précédemment, qu'ils constituent une phase d'annotation manuelle ou semi-automatique, partent pour la plupart du principe que la segmentation d'images en régions d'intérêt s'est effectuée en amont sans accroc. Cette étape n'est cependant pas triviale et certains se sont penchés sur l'intérêt que peuvent revêtir les ontologies pour le guidage et la correction des

7. parliament.semwebcentral.org/
8. www.strabon.di.uoa.gr/about
9. clarkparsia.com/pellet

algorithmes de traitement d'images.

Une ontologie peut notamment être utilisée en amont de la phase de traitement d'images afin de guider celle-ci vers l'application des opérations adaptées à un objectif donné. C'est par exemple l'approche développée par [Renouf 2007] qui, à travers l'action conjointe de deux ontologies, permet à l'expert d'un domaine d'études donné de (1) spécifier les tâches à accomplir et les objectifs à atteindre par le système de traitement d'images, (2) décrire la forme des images au niveau physique (moyens d'acquisition), perceptif (quel est le contenu visuel des images) et sémantique (que représentent ces images).

Un autre exemple consiste en un système de réédition automatique de bandes dessinées, sous un format adapté à la lecture sur un écran de téléphone, proposé par [Han 2007]. L'image d'une page est analysée afin d'extraire le texte, de séparer les cases les unes des autres et de distinguer les zones de l'image potentiellement porteuses d'informations de celles qui ne le sont pas. Les auteurs proposent alors une ontologie qui, à partir des données fournies par l'étape précédente, est à même de déduire les opérations à effectuer sur l'image (redimensionnement, troncature d'une partie de l'image jugée non importante, etc.).

[Forestier 2012, Forestier 2013] proposent une méthode d'affinement itératif de la labellisation de zones sur des prises de vue satellitaires, respectivement de zones urbaines et de zones côtières, en s'appuyant sur la cohérence des extractions avec des contraintes issues d'une connaissance du domaine modélisée a priori. C'est une approche que nous avons également explorée et qui est détaillée dans le chapitre 6.

3.6 Conclusion

La bande dessinée est un domaine n'ayant attiré l'attention de la communauté d'analyse d'images que très récemment. Les premiers travaux recensés remontent à une dizaine d'années mais la majorité des publications sont datées d'après 2010. La plupart des efforts ont jusqu'à présent été consacrés à l'exploration d'approches basées sur le traitement d'images pur. La détection de ce que l'on peut appeler *primitives* de la bande dessinée – cases, bulles, texte – a particulièrement attiré les projecteurs bien que des progrès restent encore à effectuer. La détection des personnages n'a été l'objet que de très récentes initiatives et la compréhension générale d'une planche dans sa globalité reste, à notre connaissance, un champ inexploré.

Cette dernière tâche en particulier implique de ne plus considérer les éléments indépendamment les uns des autres mais comme des parties formant un tout construit et cohérent. Une telle approche nécessite un cadre formel représentant la structure de ce tout, cadre pour la construction duquel les ontologies ont d'ores et déjà prouvé leur potentiel. Elles semblent en effet, d'une part, être adaptées à l'expression des relations sémantiques entretenues entre les éléments d'une bande dessinée, comme le prouve les initiatives récentes présentées en section 3.4. D'autre part, leur intégration dans l'étape d'analyse d'images fournit à celle-ci un contexte apportant un

bénéfice certain à la qualité des résultats d'extraction, bien que peu d'expérimentations aient été conduites sur les bandes dessinées jusqu'à présent.

Les bandes dessinées sont un type de document particulier, tant au niveau de leur contenu visuel, impliquant la création d'algorithmes de traitement d'image spécifiques, qu'au niveau de leur structure, se prêtant plutôt bien à l'exercice de la conceptualisation. Nous allons voir dans les chapitres suivants la forme que peut prendre une telle conceptualisation, réalisée dans une optique d'analyse d'images et de recherche d'informations.

Deuxième partie

Proposition d'un cadre pour l'analyse et l'interprétation automatisées d'images de bandes dessinées et son évaluation.

Chapitre 4

Présentation des ontologies proposées

Sommaire

4.1 Introduction 65
4.2 Domaine de l'image 66
4.2.1 Image et régions d'intérêt 66
4.2.2 Extracteurs 69
4.2.3 Vérité terrain 71
4.2.4 Évaluation 73
4.3 Domaine de la bande dessinée 75
4.3.1 Des albums et des planches 76
4.3.2 Contenu des planches 77
4.3.3 Spécialisation du contenu 81
4.4 Interaction entre les deux ontologies 83
4.5 Conclusion 83

4.1 Introduction

Nous présentons à partir de ce point les travaux réalisés dans le cadre de cette thèse. Ceux-ci, rappelons-le, s'inscrivent dans le cadre plus large du projet eBDtheque et se sont déroulés en parallèle de la thèse de Christophe Rigaud axée sur l'analyse de la bande dessinée d'un point de vue traitement d'images.

Nous introduisons dans ce chapitre les ontologies développées pour l'analyse automatique de bandes dessinées. Nous progressons en plusieurs étapes. Tout d'abord, nous présentons une première ontologie formalisant les notions relatives au traitement d'images, le type de données, les entrées, les sorties et l'évaluation des résultats. Dans un deuxième temps, nous déroulons le processus de conceptualisation du domaine de la bande dessinée, pensée pour être utilisée dans un contexte d'analyse d'images. Enfin, nous expliquons comment ces ontologies communiquent l'une avec l'autre, afin de pouvoir être utilisées conjointement.

La mise en œuvre de ces ontologies est illustrée, à travers plusieurs exemples d'application, dans la troisième partie de ce document.

4.2 Domaine de l'image

4.2.1 Image et régions d'intérêt

Nous avons développé dans un premier temps un modèle formalisant les notions primitives du traitement d'images. Ce modèle se veut volontairement générique en n'intégrant aucun concept relatif à la bande dessinée afin de pouvoir être exploité dans d'autres domaines relatifs à l'analyse de documents. Il sert de support à la structuration des données issues d'algorithmes de traitement d'images en vue de leur enrichissement sémantique ultérieur. Les données produites par de tels algorithmes se réduisent généralement à des données spatiales, des lignes, des régions, auxquelles nous ferons indifféremment référence sous les notions de *régions d'intérêt* ou *ROI* (pour Region Of Interest) dans la suite de ce document. Elle sont définies par leurs coordonnées cartésiennes dans le repère orthonormé de l'image desquelles elles sont extraites. En partant de cette vision à gros grains de ce qu'est le traitement d'image, nous définissons les deux premiers concepts de notre modèle : *Image* et *ROI*.

Le concept *Image* modélise la notion d'image en tant qu'objet numérique, en tant que matière première du système de traitement d'images. C'est le concept le plus général à partir duquel seront dérivés les autres concepts de notre modèle. Un algorithme de traitement d'images fonctionne en manipulant directement les pixels de l'image, ses composants élémentaires. Il peut les analyser indépendamment les uns des autres ou bien les regrouper par ensembles mais au final, la valeur, l'intensité de chaque pixel a son importance dans le processus d'analyse et peut avoir des conséquences sur les résultats de sortie. Or, pour deux images d'apparence similaire aux yeux d'un observateur humain, la valeur individuelle de chacun de leurs pixels peut grandement varier d'une image à l'autre. Elles peuvent avoir une définition différente, c'est à dire des dimensions (mesurées en pixels) différentes. Dans le cas d'images numérisées, si les documents d'origine étaient de même taille, cette différence de définition est la conséquence d'une différence de résolution de numérisation. La résolution d'une image, indiquée par l'attribut *hasResolution*, correspond au nombre de pixels utilisés pour discrétiser un pouce de l'espace. Plus la résolution est élevée, plus l'image est nette. De plus, l'encodage de l'information visuelle au sein de chaque pixel varie également en fonction du format d'image utilisé et du degré de compression avec ou sans pertes appliqué. Il est important d'intégrer au modèle ces informations à propos de chaque image car elles sont le contexte de travail dans lequel se place l'outil d'analyse. Elles conditionnent et permettent d'interpréter et de comprendre la qualité des résultats produits. La notion de définition est dérivée en deux attributs, *hasWidth* et *hasHeight*, prenant pour valeurs des entiers. Associés à la résolution, ils permettent au besoin de retrouver la taille réelle de l'image d'origine et couvrent ainsi l'intégralité des informations relatives aux conditions de numérisation. Le format de l'image, ainsi que l'information binaire de compression avec ou sans perte, sont respectivement représentés par les attributs *hasFormat* et *isLossLess*. Le format de l'image est indiqué par une chaîne de caractères, par exemple

"PNG" ou "TIFF".

Dans le domaine de l'analyse d'image, les fichiers picturaux numériques sont manipulés d'une manière semblable à des tableaux à deux dimensions, éventuellement trois si l'on considère que la décomposition de l'information de couleur en plusieurs canaux distincts nécessite une dimension supplémentaire. Dès lors, l'origine du repère de l'image correspondra classiquement à son coin supérieur gauche, comme l'est naturellement celle d'un tableau, plutôt qu'à son coin inférieur gauche, comme l'est traditionnellement celle d'un graphique ou de la représentation d'une fonction mathématique. Nous adoptons cette convention dans notre modèle et considérons que le pixel situé dans le coin supérieur gauche d'une image de largeur l et de hauteur h a pour coordonnées $(x, y) = (0, 0)$ et que celui situé dans le coin inférieur droit a pour coordonnées $(x, y) = (l, h)$.

Le concept *ROI* est un méronyme du concept *Image*, c'est à dire qu'une région d'intérêt est une portion d'image, qu'elle fait partie d'une image. Ce concept formalise la notion de région d'intérêt telle qu'elle est perçue par l'analyste d'images, c'est à dire une surface finie incluse dans le plan de l'image, composée de pixels connexes possédant des caractéristiques visuelles correspondant à une certaine classe d'objets visuels recherchés. Le fait qu'une région d'intérêt soit une surface en deux dimensions, finie et fermée, nous permet de la représenter sous la forme d'un polygone.

Les coordonnées du polygone sont stockées dans la A-Box de notre modèle sous la forme d'une chaîne de caractères et exprimées au format Well-Known Text, souvent abrégé WKT. Ce format est une initiative de l'Open Geospatial Consortium (ou OGC) [Open GIS Consortium Inc. 2011] qui travaille, conjointement avec l'ISO/TC 211, à le faire évoluer vers un de ses standards ainsi qu'à en faire une norme ISO internationale. Nous avons fait le choix de ce formalisme parmi d'autres, tels que SVG [Ferraiolo 2001] ou GeoJSON [Butler 2008], pour plusieurs raisons. Premièrement, et comme la plupart des langages, il permet d'exprimer un grand nombre de formes géométriques différentes dont les polygones et les lignes qui nous intéressent dès à présent. Au delà de ces formes de base, il permet également de représenter des polygones contenant des trous ainsi que des multi-polygones, c'est à dire un ensemble de polygones, potentiellement disjoints et induisant donc une perte de connexité de la région d'intérêt, mais constitutifs de la même entité. Deux capacités qui montreront leur intérêt en perspective de ces travaux. Il semble donc à la fois assez expressif pour couvrir des besoins de représentation de régions issues de l'analyse d'un document, tout en évitant d'intégrer des notions non pertinentes pour notre contexte d'étude qui rendrait la syntaxe trop lourde. Il présente en effet l'avantage, non négligeable lorsque l'on est amené à traiter de grands volumes de données, d'être concis, sans pour autant réduire la nature explicite de son formatage. Celui-ci permet à un utilisateur humain d'identifier en un coup d'œil la nature de l'objet géométrique décrit. Cette nature est exprimée en début de séquence sous la forme d'un mot clé, tel que `POINT`, `LINESTRING` ou `POLYGON`, respectivement pour

les points, les lignes et les polygones. Le listing A.1 met en évidence le coté concis du format WKT en comparant sa manière de décrire un pentagone avec celle des formats SVG, GeoJSON et Geography Markup Language.

```
Well-Known Text :
  POLYGON((0 0, 50 0, 50 100, 0 100, 25 50, 0 0))

SVG :
  <svg>
    <polygon points="0,0 50,0 50,100 0,100 25,50 0,0"/>
  </svg>

GeoJSON :
{
  "type": "Polygon",
  "coordinates": [
    [[0, 0], [50, 0], [50, 100], [0, 100], [25, 50], [0, 0]]
  ]
}

Geography Markup Language :
  <gml:Polygon>
    <gml:exterior>
      <gml:LinearRing>
        <gml:coordinates>0,0 50,0 50,100 0,100 25,50 0,0</gml:coordinates>
      </gml:LinearRing>
    </gml:exterior>
  </gml:Polygon>
```

ENCADRE 4.1 – Expression d'un polygone dans différents formats

Enfin, le format WKT a l'avantage d'être un format ouvert, bien documenté et supporté dans de nombreuses API développées pour des langages de programmation, allant du C au Javascript, en passant par Python, Java et .NET, aux usages bien différents. Un critère ne devant pas être négligé lorsque l'on souhaite proposer des outils réutilisables dans différents contextes.

Chaque région d'intérêt possède alors un attribut *hasWKT*, prenant pour valeur une chaîne de caractères, permettant de décrire à la fois son type (un polygone, une ligne, etc.), et ses coordonnées. Les attributs additionnels *hasArea*, *hasX* et *hasY* permettent respectivement d'exprimer l'aire de la région d'intérêt et les coordonnées de son centre.

La relation *hasROI* est définie afin de lier les concepts *Image* et *ROI*. Elle a pour origine une instance du premier et pour cible une instance du second. Elle symbolise le fait qu'une image peut contenir une ou plusieurs régions d'intérêt. En revanche, une région d'intérêt fait partie d'une et d'une seule image. Une ROI est en effet issue d'un travail d'analyse d'une image donnée ayant pour but la délimitation d'une zone ayant un intérêt pour une application particulière. La même analyse conduite

sur des images aux caractéristiques différentes produira immanquablement des régions d'intérêt aux caractéristiques différentes. La propriété *hasROI* est alors notée comme étant inversement fonctionnelle afin d'intégrer cette contrainte. La figure 4.1 synthétise les notions exposées jusqu'à présent.

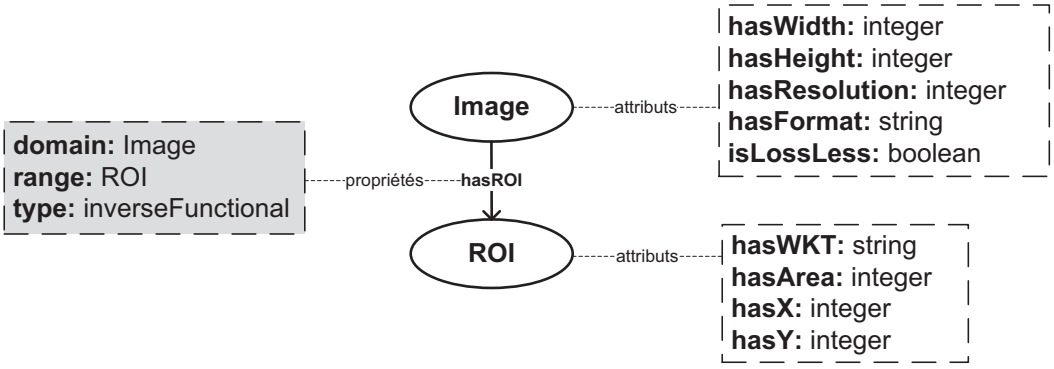


FIGURE 4.1 – Les concepts *Image* et *ROI*, leurs relations et leurs attributs.

4.2.2 Extracteurs

Le traitement d'images, et particulièrement la reconnaissance de formes, est une discipline complexe. L'algorithme générique permettant d'identifier n'importe quel type de forme, n'importe quel objet dans une image, qu'elle représente une scène naturelle, un document ou qu'elle ait été générée numériquement, n'existe pas. En fonction de la tâche qu'il cherche à accomplir, du type de contenu qu'il souhaite détecter, voire reconnaître, dans une image, le chercheur met en place des techniques adaptées aux caractéristiques visuelles de ce contenu. Il va par exemple tantôt se concentrer sur la texture, tantôt sur la couleur selon qu'il souhaite identifier des formes aux motifs connus et répétitifs, comme la surface de la mer ou le pantalon blanc et bleu d'Obélix, ou identifiables par une teinte donnée, comme par exemple un ciel bleu ou le pantalon rouge d'Astérix. La forme de ces éléments, leur taille, leur position absolue ou relative ou encore leur orientation sont autant d'éléments parmi bien d'autres permettant de façonner un outil adapté à la tâche à accomplir.

Un algorithme de traitement d'images a donc généralement pour fonction la détection d'un seul type d'objet. Toutefois, si son domaine de compétence est limité, il est souhaitable que cet algorithme soit assez robuste pour produire des résultats corrects sur un grand nombre d'images différentes. Dès lors, le chercheur souhaitant analyser un ensemble de documents complexes mais de même nature, dans le sens où ceux-ci sont composés d'un ensemble d'éléments graphiques différents mais toutefois communs au corpus de documents, va devoir mettre en place autant de traitements spécifiques que de types d'éléments qu'il souhaite extraire, tout en prenant garde à

ce qu'ils fonctionnent sur le plus grand nombre de documents possible.

Si nous nous plaçons par exemple dans le cadre d'un traitement automatique de documents administratifs, tels que des fiches de paye ou des factures, il pourrait être intéressant de mettre en place des systèmes permettant de détecter les logos, les caractères alphanumériques ou encore les signatures. Si nous nous replaçons dans le contexte de la bande dessinée, la segmentation des cases, des bulles, du texte et des personnages, nécessite autant d'algorithmes d'analyse d'images différents, comme nous avons pu le voir en section 3.3. Chaque algorithme va tirer parti de ce qui fait la spécificité des éléments à extraire. Les cases sont des éléments majeurs de l'image en terme d'espace occupé dans la page. Elles sont généralement délimitées par une bordure noire et séparées les unes des autres par une gouttière. Leur forme, souvent rectangulaire, la plupart du temps au moins polygonale est également un critère utilisé, notamment via la détection des lignes dans l'image. Les bulles contiennent du texte, traditionnellement de couleur foncée sur un fond clair. Elles sont également délimitées par un contour noir mais ont généralement des dimensions inférieures à celles de cases. Les lignes de texte sont composées de petits éléments visuels, contrastant avec le fond de l'image, ponctuellement alignés de manière régulière. Des techniques de reconnaissance de caractères (OCR) peuvent éventuellement venir en renfort afin de filtrer les faux-positifs. Les personnages de bande dessinée peuvent avoir n'importe quelle forme, réaliste ou caricaturale, humanoïde ou non. Ils présentent en revanche une certaine régularité dans leur représentation visuelle. Ils peuvent changer de position, d'orientation ou de dimensions mais ils conservent la plupart du temps les mêmes attributs vestimentaires, les mêmes caractéristiques visuelles, au moins le long d'une même histoire.

Une région d'intérêt identifiée par un algorithme de segmentation d'images correspond donc à la proposition par ce dernier d'une zone spatiale de l'image contenant le type d'élément qu'il a été conçu pour détecter. Cette région peut effectivement contenir un élément de ce type ou bien contenir tout autre chose. La précision de segmentation de l'algorithme est dépendante à la fois des caractéristiques de l'image analysée, de la complexité du domaine d'analyse et bien sûr de la qualité de son implémentation. L'ensemble des résultats peut contenir des erreurs. Les régions d'intérêt produites ne sont que des propositions demandant à être vérifiées, elles n'ont pas valeur de vérité.

Nous appellerons de tels systèmes de segmentation *extracteurs* dans la suite de ce document. La notion d'extracteur est intégrée dans notre modèle sous le concept *Extractor*. Relié à celui de *ROI*, il permet d'exprimer le lien étroit entre l'algorithme de traitement d'images et les régions d'intérêt qu'il produit à partir d'une image aux caractéristiques données. Cet ensemble de notions fournit un premier cadre à l'évaluation de la pertinence de l'intégration d'une ROI dans un système plus complet où elle sera amenée à interagir avec d'autres entités. Une région d'intérêt provenant d'une image est produite par un et un seul extracteur. C'est pourquoi la relation

hasExtractor liant les concepts *ROI* et *Extractor* est définie comme propriété fonctionnelle. Le type d'objets recherché par l'extracteur peut être indiqué à travers l'attribut *hasROIType*. L'ensemble des types possibles devant être défini en fonction du contexte d'application.

La figure 4.2 illustre ces ajouts au modèle.

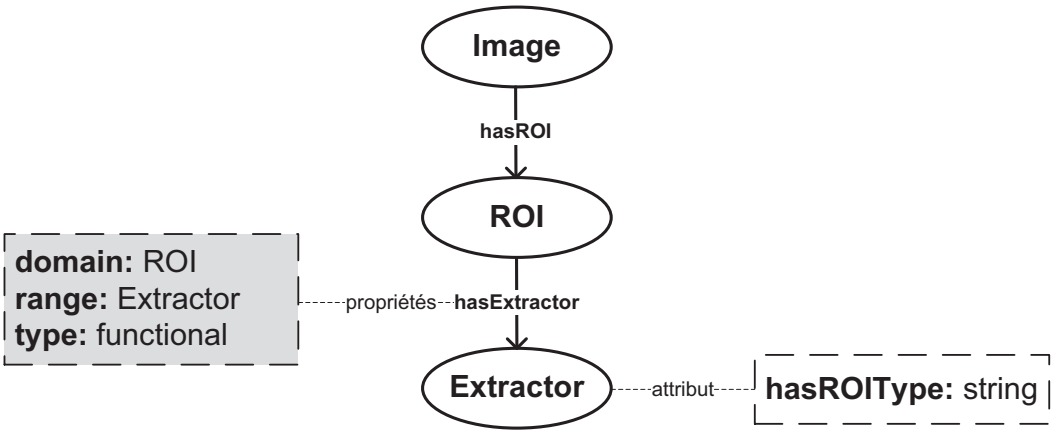


FIGURE 4.2 – Modèle présenté en figure 4.1 enrichi de la notion d'extracteur.

4.2.3 Vérité terrain

A la notion de détection ou de reconnaissance de formes et d'objets s'associe invariablement celle de son évaluation. Comme nous l'avons évoqué, les résultats produits par de telles méthodes ne sont que des propositions, conséquences brutes des caractéristiques visuelles des pixels d'une image passés au crible des instructions d'un algorithme d'extraction. La proportion de *faux positifs* (propositions ne correspondant pas à des éléments à extraire) et de *faux négatifs* (éléments non identifiés) dans les résultats varie d'un domaine, d'une image et d'un algorithme à l'autre, mais n'est jamais, sinon très rarement, nulle. On parle notamment de mesures de précision et de rappel. Ces deux mesures s'expriment par un nombre compris entre zéro et un. Soit un ensemble d'éléments à détecter S et un ensemble d'éléments détectés S'. La précision correspond à la proportion d'éléments correctement détectés parmi tous les éléments détectés, formellement $|S \cap S'|/|S'|$. Elle mesure la tendance qu'a un algorithme à sur-détecter, ou non, les éléments dans un jeu de données. On peut vouloir, selon les applications, assurer la précision des résultats, quitte à omettre certains éléments aux dépends de l'exhaustivité. Les capacités d'exhaustivité d'un algorithme se mesurent grâce au rappel. Le rappel correspond à la proportion d'éléments correctement détectés parmi tous les éléments qu'il y avait à détecter dans le corpus, formellement $|S \cap S'|/|S|$. On aura tendance à maximiser le rappel dans les cas où la présence d'un grand nombre de faux-positifs n'est pas un critère de per-

turbation et où le fait de détecter un maximum d'éléments du corpus prime avant toute chose. Avant de pouvoir utiliser directement ces résultats dans des systèmes informatiques à destination d'utilisateurs humains réels, le scientifique doit s'assurer de la robustesse de ses méthodes, notamment à travers ces mesures de précision et de rappel. Lorsqu'elles sont calculées sur un corpus de documents représentatif du domaine d'application, elles fournissent l'assurance que les méthodes produisent des résultats en concordance avec les critères de qualité exigés pour une application donnée.

Le calcul de ces mesures nécessite un ensemble d'informations, connu a priori, portant sur le contenu des documents. Ces informations sont la référence sur laquelle va s'appuyer le chercheur pour valider ses résultats et elles ne seront jamais remises en cause. Elles sont absolues, définissent ce qui est vrai et tout ce qui est vrai dans une logique de monde fermé où, par conséquent, tout ce qui n'est pas explicitement défini n'existe pas. On nomme cet ensemble d'informations *vérité terrain*. Une vérité terrain représente la réalité d'un document, ou même plus généralement de n'importe quel type de contenu, sous la forme d'annotations le décrivant parfaitement et complètement pour un contexte applicatif donné. En effet, selon les résultats que l'on cherche à produire à partir de l'analyse d'un ensemble de documents, la référence nécessaire à leur évaluation ne sera pas nécessairement de même nature ni de même granularité. Dans le cas par exemple où l'on cherche à déterminer la nature d'un document (photographie, article de presse, page de bande dessinée, etc.), un simple mot clé spécifiant cette nature peut constituer une vérité terrain suffisante. Si l'on souhaite en revanche analyser son contenu, sa structure interne, une annotation plus profonde doit être effectuée. Pour l'analyse du contenu d'une page de bande dessinée, cela passe notamment par la segmentation spatiale des éléments que l'on souhaite extraire, par exemple les cases ou les personnages. Des annotations sémantiques peuvent également être associées à ces éléments. Le critère de vérité absolue intrinsèque à une vérité terrain implique que celle-ci ne soit pas générée automatiquement mais construite manuellement par un humain, expert du domaine. En fonction de la granularité des annotations, la construction d'une vérité terrain peut donc se révéler très coûteuse en temps. La connaissance de l'expert formalisée par l'ensemble de ses annotations représente l'étalon d'or, la limite indépassable, dont doit s'approcher l'algorithme de traitement automatique pour prétendre à un fonctionnement proche de la compréhension humaine d'un document. Si dans certains domaines la subjectivité de l'expert n'influence pas l'annotation, il en existe, c'est notamment le cas de la bande dessinée et de nombreux autres domaines où la liberté artistique est de mise, où la vérité terrain sera dépendante de l'interprétation personnelle de l'annotateur. C'est pour cela que l'on parle d'*une* et non de *la* vérité terrain d'un corpus de documents pour un domaine d'application. Nous avons construit une telle vérité terrain pour la bande dessinée. Le chapitre 5 est consacré à sa description.

Si l'on se place dans le contexte de l'analyse du contenu des images, où des traitements spécifiques extraient des régions spatiales candidates pour une classe

d'éléments donnée, la vérité terrain correspondante sera également composée d'un ensemble de régions spatiales, annotées de la classe à laquelle elles appartiennent. Si l'on écarte un instant la méthode de production de ces régions d'intérêt, une vérité terrain correspond à la définition d'extracteur posée précédemment. Nous pouvons donc formaliser ce concept dans notre modèle par une relation d'hyponymie entre la classe *Extractor* et deux spécialisations *GroundTruth* et *ExtractorAuto*, modélisant respectivement les notions de vérité terrain et d'extraction automatique. Ces deux concepts sont cependant fondamentalement disjoints car les extracteurs qu'ils définissent produisent des régions d'intérêt aux fonctions différentes. Celles issues d'une segmentation automatique sont des propositions qui, à terme, viendront peupler un système informatique destiné à être utilisé. Elles peuvent être remises en cause, être manipulées, modifiées ou supprimées. Les ROI provenant d'une vérité terrain sont en revanche, de part leur statut de référence absolue, intouchables. Elles ne servent qu'à valider ponctuellement les propositions d'un algorithme et n'ont pas vocation à intégrer un quelconque système en production. Cette différence de nature est modélisée par la spécialisation du concept *ROI* en deux concepts *ROIGT*, *GT* pour Ground Truth, et *ROIAuto*. Ces deux concepts, comme les deux extensions d'*Extractor*, sont disjoints, c'est à dire qu'une région ne peut pas être à la fois proposition et référence. Un extracteur ne peut pas non plus être automatique et en même temps avoir valeur de vérité. Le concept *ROIGT*, respectivement *ROIAuto*, est noté équivalent à l'intersection des ensembles d'individus de la classe *ROI* et de l'ensemble des individus ayant pour extracteur une instance du concept *GroundTruth*, respectivement *ExtractorAuto*. Cette équivalence est transcrite en syntaxe de Manchester [Horridge 2006] dans équation 4.1.

$$\text{ROIGT} \equiv \text{ROI } \textbf{and } \text{hasExtractor } \textbf{some } \text{GroundTruth} \tag{4.1}$$

$$\text{ROIAuto} \equiv \text{ROI } \textbf{and } \text{hasExtractor } \textbf{some } \text{ExtractorAuto} \tag{4.2}$$

La figure 4.3 illustre l'intégration de ces notions dans le modèle.

4.2.4 Évaluation

Une fois la notion de référence conceptualisée, il est nécessaire de mettre en place la structure permettant d'évaluer les régions segmentées par rapport à cette référence. En analyse d'images, il est imprudent d'espérer d'un algorithme une annotation des régions extraites permettant d'identifier individuellement ces dernières et, dès lors, de les comparer à leur référence dans la vérité terrain. Les problèmes sont souvent trop nombreux pour cela, le nombre de régions extraites ne correspondant pas nécessairement au nombre de régions à détecter et l'algorithme devant travailler sans connaissance a priori du contenu des instances. Les seules informations fournies par un extracteur sur lesquelles il est possible de fonder une réflexion sont le type

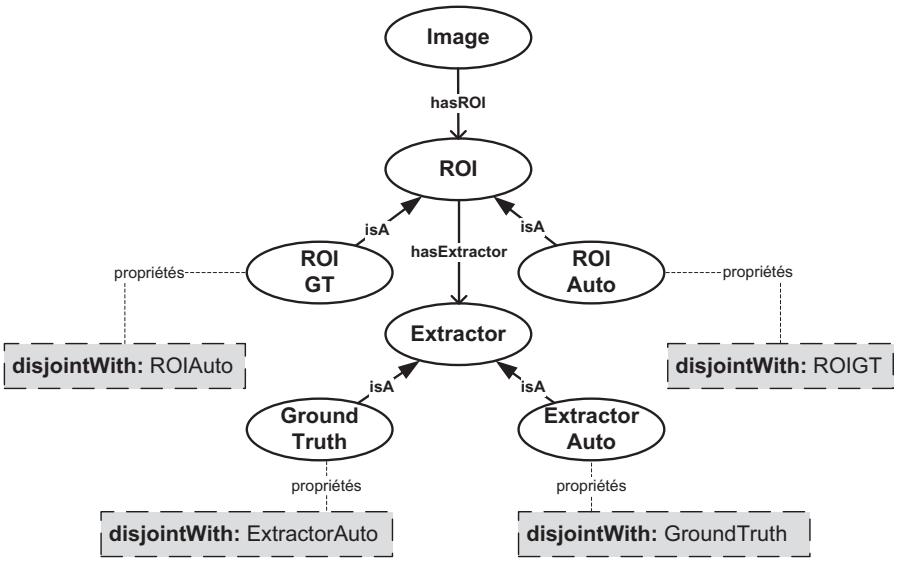


FIGURE 4.3 – Modèle présenté en figure 4.2 enrichi de la notion de vérité terrain.

de la région extraite (la fonction de l'algorithme) et la localisation de cette région dans l'image. Dès lors, afin de valider les régions extraites, il est nécessaire de les mettre en relation avec chaque région de la vérité terrain pour lesquelles elles partagent un ensemble non-vide de pixels. Plus la cardinalité de l'intersection des deux éléments est proche de la cardinalité de leur union, meilleure est la qualité de la détection. Cette qualité peut être quantifiée par les mesures de rappel et de précision introduites en sous-section 4.2.3, les éléments à détecter étant ici les pixels. Ces mesures peuvent être combinées dans le calcul du F-score afin de produire une valeur unique représentative de la qualité générale de l'extraction. Le F-score classique est égal $2.(P \times R)/(P + R)$ mais il peut éventuellement être pondéré pour donner plus de poids à l'une ou l'autre des parties. En fonction des contraintes de l'application, on peut être amené à vouloir trancher de la validité ou non d'une région via l'introduction d'un seuillage. Une région d'intérêt sera alors considérée comme une extraction valide si, par exemple, sa valeur de F-Score dépasse un certain seuil s.

La notion de correspondance entre une région à évaluer et une région de référence est formalisée par le concept *Evaluation*. Celui-ci est complété par trois attributs numériques, *hasRecall*, *hasPrecision*, *hasFScore* permettant d'associer respectivement à ses instances des valeurs de rappel, précision et un F-score. Les relations *hasReference* et *hasCandidate* permettent respectivement de lier des instances de ce concept à des instances de *ROIGT* et *ROIAuto*. Ces deux relations sont fonctionnelles, chaque évaluation ne pouvant impliquer qu'une seule région candidate et qu'une seule référence. Une instance d'évaluation peut être créée pour chaque intersection spatiale entre une région d'intérêt provenant d'une vérité terrain et une région d'intérêt provenant d'un extracteur automatique. Les relations retenues à cet

effet sont la disjonction (*disjoint*), l'inclusion (*contains*) et l'intersection (*overlaps*) mais l'intégralité des relations du RCC-8 est compatible via l'importation des ontologies spatiales définies pour Pellet Spatial [1] et GeoSPARQL [2]. L'équivalence entre le concepts *ROI* et le concept *SpatialObject*, formalisé par cette dernière, permettant l'usage de ses propres relations spatiales. Les mesures de rappel, précision et F-score sont calculées et insérées comme attribut de l'instance d'évaluation.

Les concepts *Validation* et *Error* sont introduits comme des classifications additionnelles pour les instances d'*Evaluation*. Ils sont définis comme équivalents à l'ensemble des instances d'*Evaluation* ayant un attribut *hasFScore* d'une valeur supérieure (respectivement, inférieure) à un seuil s donné (voir équation 4.3). Cette règle peut être modifiée en fonction des besoins pour se baser uniquement sur le rappel ou la précision, voire sur d'autres métriques telles que celle introduite par VOC [Everingham 2009]. Ils sont de plus considérés comme disjoints, une évaluation ne pouvant pas être à la fois valide et invalide.

$$\text{Validation} \equiv \text{Evaluation } \textbf{and } \text{hasFScore } \textbf{some } \text{double}[>=s] \tag{4.3}$$

$$\text{Error} \equiv \text{Evaluation } \textbf{and } \text{hasFScore } \textbf{some } \text{double}[<s] \tag{4.4}$$

Le choix de l'ajout d'un concept spécifique pour représenter la notion d'évaluation, plutôt qu'un attribut sur *ROI* ou une propriété entre *ROIGT* et *ROIAuto*, se justifie lors d'évaluations multiples. Notamment lorsqu'une région possède une intersection avec plusieurs références ou lorsque plusieurs vérités terrain sont insérées dans la A-Box. Dans ces cas de figure, l'utilisation d'attributs et de relations directes entre régions d'intérêt pour représenter la validité par rapport à une référence n'est plus suffisante.

La figure 4.4 illustre ces ajouts au modèle.

4.3 Domaine de la bande dessinée

L'ontologie présentée en section 4.2 permet d'organiser les données brutes produites en sortie d'un système d'analyse automatique d'images dans une structure formelle. Il permet de mettre en correspondance ces données avec des bases de références et d'en proposer une évaluation simple et immédiate. Nous présentons dans cette section une conceptualisation du domaine de la bande dessinée ainsi que sa formalisation ontologique. Cette conceptualisation a été pensée en gardant à l'esprit l'objectif d'une utilisation dans le cadre de l'analyse d'images. Des amendements y seront apposés plus tard dans ce manuscrit afin de rendre son utilisation possible dans d'autres cadres applicatifs.

1. URI : http ://clarkparsia.com/pellet/spatial#
2. URI : http ://www.opengis.net/ont/geosparql#

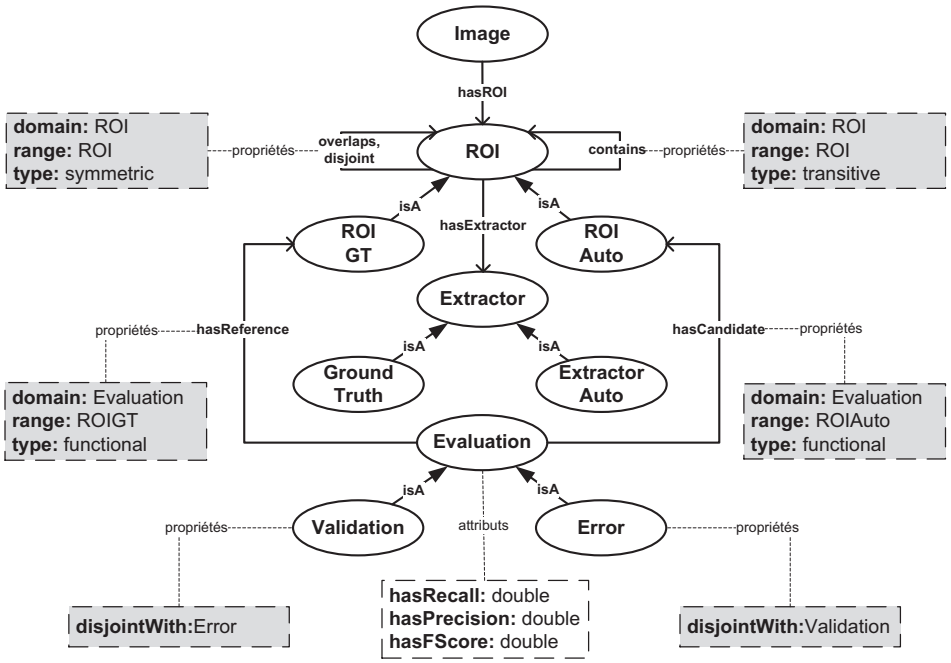


FIGURE 4.4 – Modèle présenté en figure 4.3 enrichi de la notion d'évaluation, de validation et d'erreur.

4.3.1 Des albums et des planches

Comme nous l'avons vu dans le chapitre 1, une bande dessinée se définit comme une suite d'images transmettant un message. Ces images sont spatialement juxtaposées dans un plan que l'on nomme *planche*. Dans la bande dessinée que l'on peut qualifier de classique, franco-belge, américaine et japonaise, une histoire se raconte à travers une succession ordonnée de planches. Celles-ci sont matérialisées par des pages imprimées regroupées au sein d'un album, ce dernier pouvant lui même n'être qu'une itération d'une collection contant une histoire plus large. Les webcomics font également usage de planches, matérialisées par des images numériques. Dans le cas de la bande dessinée imprimée, une planche peut s'étendre sur une ou deux pages, rarement plus. Une planche de webcomic est toujours représentée par une image unique.

Les deux premiers concepts introduits sont alors *Comic* et *Plate*[3], représentant respectivement les notions d'album et de planche. Ces deux concepts sont liés par la propriété *hasPlate*. Celle-ci est inversement fonctionnelle, une planche ne pouvant faire partie que d'un seul album.

Les informations bibliographiques de chaque album sont représentées par des attributs associés au concept *Comic*. Le titre de l'album, l'éventuelle collection de

3. Le vocabulaire français possède une distinction claire entre les termes planche et page. Les anglophones ont cependant tendance à utiliser le mot *page* pour désigner ces deux notions. Le terme *plate*, moins usité, a été sélectionné afin de désambiguïser ces deux notions.

laquelle il est issu, ses auteurs, sa date de publication ou encore son ISBN peuvent être renseignés via les attributs correspondants. La direction de lecture (de gauche à droite ou de droite à gauche) peut être indiquée à travers l'attribut booléen *right2Left*.

Les attributs de *Plate*, *hasNumber* et *onDoublePage*, permettent respectivement d'indiquer le numéro d'une planche et le nombre de pages qu'elle occupe. La succession des planches est formalisée par la propriété *hasNextPlate* dont l'instanciation est dérivée de l'attribut *hasNumber* par la règle SWRL exprimée en équation 4.5.

$$\begin{aligned} &\text{Comic}(\,?c), \text{Plate}(\,?p1), \text{Plate}(\,?p2),\\ &\text{hasPlate}(\,?c, ?p1), \text{hasPlate}(\,?c, ?p2),\\ &\text{hasNumber}(\,?p1, ?n1), \text{hasNumber}(\,?p2, ?n2),\\ &\qquad \text{subtract}(1, ?n2, ?n1) \rightarrow \text{hasNextPlate}(\,?p1, ?p2) \end{aligned} \tag{4.5}$$

La figure 4.5 illustre ces premiers éléments.

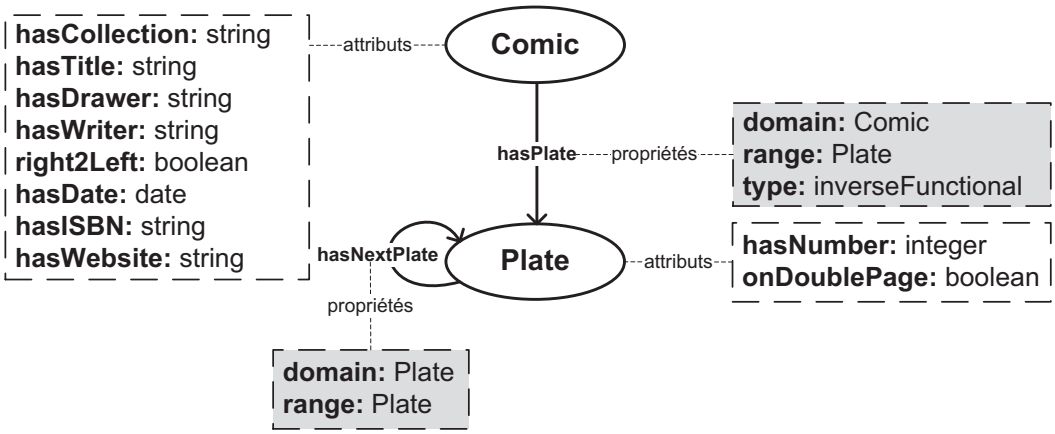


FIGURE 4.5 – Les concepts *Comic* et *Plate*, leurs relations et leurs attributs.

4.3.2 Contenu des planches

Comme présenté dans le chapitre 1, une planche de bande dessinée est le cadre dans lequel sont placées les cases constituant le récit. L'ordre de ces cases dans la séquence de lecture est défini par leur position relative les unes par rapport aux autres ainsi que, dans certains cas, l'interprétation personnelle du lecteur. L'étude et la génération de la séquence de lecture est l'objet de la section 7.2. Le contenu des cases est composé de dessins représentant notamment des personnages et des phylactères. Nous avons choisi de nous concentrer sur les relations entre ces deux

types de contenu dans un premier temps. La gestion des autres éléments représentés au sein des cases dépasse le cadre de ces travaux, bien qu'une approche basée sur une extraction de termes appropriés depuis WordNet, similaire à celle proposée par [Zinger 2005], pourrait se révéler intéressante. Les phylactères contiennent eux-même des lignes de texte, matérialisant les mots des personnages et la narration de l'histoire.

Les cases sont représentées par le concept *Panel*. Celui-ci possède un attribut *hasRank* indiquant la place de chacune de ses instances dans la séquence de lecture de la planche à laquelle elles sont rattachées. Une propriété *hasNextPanel* lie chaque case à son successeur dans la page. Les instances de ces propriétés sont générées via l'application de la règle 4.6 exprimée en SWRL. La propriété *hasNextPanel* n'est pas fonctionnelle car cette séquence n'est pas nécessairement unique. En effet, certaines mises en page introduisent des degrés de liberté dans le parcours des planches, comme nous le verrons en section 7.2.

$$\begin{aligned} \text{Plate}(?p),\ \text{Extractor}(?e),\\ \text{Panel}(?p1),\ \text{Panel}(?p2),\\ \text{hasPanel}(?p, ?p1),\ \text{hasPanel}(?p, ?p2),\\ \text{hasExtractor}(?p1, ?e),\ \text{hasExtractor}(?p2, ?e),\\ \text{hasRank}(?p1, ?r1),\ \text{hasRank}(?p2, ?r2),\\ \text{subtract}(1, ?r2, ?r1) \rightarrow \text{hasNextPanel}(?p1, ?p2) \end{aligned} \tag{4.6}$$

Les phylactères, qu'il s'agisse de bulles de dialogue, de pensée ou de cadres de narration, sont représentés par le concept *Balloon*. D'une manière similaire aux cases dans la planche, les bulles doivent être lues selon un ordre défini par leur position dans leur case de rattachement. On parle de case de *rattachement* et non pas de case *englobante* car, selon les styles de bandes dessinées, les bulles se sont pas nécessairement positionnées spatialement à l'intérieur d'une case. Il arrive qu'elles soient légèrement à l'extérieur ou à cheval sur plusieurs cases. En revanche, chaque case met en images une action se déroulant à un instant fixe. Les bulles font partie intégrante de la mise en scène et sont donc rattachées à une case donnée, quelle que soit leur position par rapport à cette case. Leur place dans l'ordre de lecture d'une case est défini par l'attribut *hasRank* et la propriété *hasNextBalloon* lie les bulles entre elles selon la formule 4.7.

$$\begin{aligned} &\text{Panel}(\,?\text{p}),\ \text{Extractor}(\,?\text{e}),\\ &\text{Balloon}(\,?\text{b1}),\ \text{Balloon}(\,?\text{b2}),\\ &\text{hasBalloon}(\,?\text{p},?\text{b1}),\ \text{hasBalloon}(\,?\text{p},?\text{b2}),\\ &\text{hasExtractor}(\,?\text{b1},?\text{e}),\ \text{hasExtractor}(\,?\text{b2},?\text{e}),\\ &\text{hasRank}(\,?\text{b1},?\text{r1}),\ \text{hasRank}(\,?\text{b2},?\text{r2}),\\ &\text{subtract}(1,?\text{r2},?\text{r1}) \rightarrow \text{hasNextBalloon}(\,?\text{b1},?\text{b2}) \end{aligned} \tag{4.7}$$

Les bulles de dialogue possèdent une ou plusieurs excroissances sur leur contour, appelées flèches ou queues, pointant vers le ou les personnages dont elles matérialisent les paroles. Cette flèche est représentée par le concept *Tail*, et sa direction par l'attribut *hasDirection*. Une bulle est liée à une ou plusieurs flèches par la propriété *hasTail*.

Les lignes de texte sont représentées par le concept *TextLine*. Elles sont regroupées à l'intérieur de phylactères et se lisent naturellement de haut en bas. Comme pour les cases et les bulles, l'attribut *hasRank* permet de spécifier leur place dans la séquence de lecture au sein d'une bulle, la propriété *hasNextTextLine* étant définie par la formule 4.8. Le texte de la ligne est retranscrit par l'attribut *hasText*.

$$\begin{aligned} &\text{Balloon}(\,?\text{b}),\ \text{Extractor}(\,?\text{e}),\\ &\text{TextLine}(\,?\text{t1}),\ \text{TextLine}(\,?\text{t2}),\\ &\text{hasTextLine}(\,?\text{b},?\text{t1}),\ \text{hasTextLine}(\,?\text{b},?\text{t2}),\\ &\text{hasExtractor}(\,?\text{t1},?\text{e}),\ \text{hasExtractor}(\,?\text{t2},?\text{e}),\\ &\text{hasRank}(\,?\text{t1},?\text{r1}),\ \text{hasRank}(\,?\text{t2},?\text{r2}),\\ &\text{subtract}(1,?\text{r2},?\text{r1}) \rightarrow \text{hasNextTextLine}(\,?\text{t1},?\text{t2}) \end{aligned} \tag{4.8}$$

Enfin, les personnages sont représentés par le concept *Character*. Les concepts *Panel*, *Balloon*, *Tail*, *TextLine* et *Character* sont disjoints, un élément ne pouvant être instance que de l'un d'entre eux.

La figure 4.6 illustre l'ajout de ces concepts à notre ontologie.

Plusieurs propriétés sont introduites dans notre ontologie afin de représenter les liens existants entre les différents éléments composant une planche. Une case étant relative à une planche, la propriété *hasPanel* lie une instance de *Plate* à une instance de *Panel*. Le *domain* et le *range* de la propriété assurent que seule une planche et une case puisse en être respectivement le sujet et l'objet. La propriété *hasPanel* est également définie inversement fonctionnelle, une case ne pouvant faire partie que d'une seule planche.

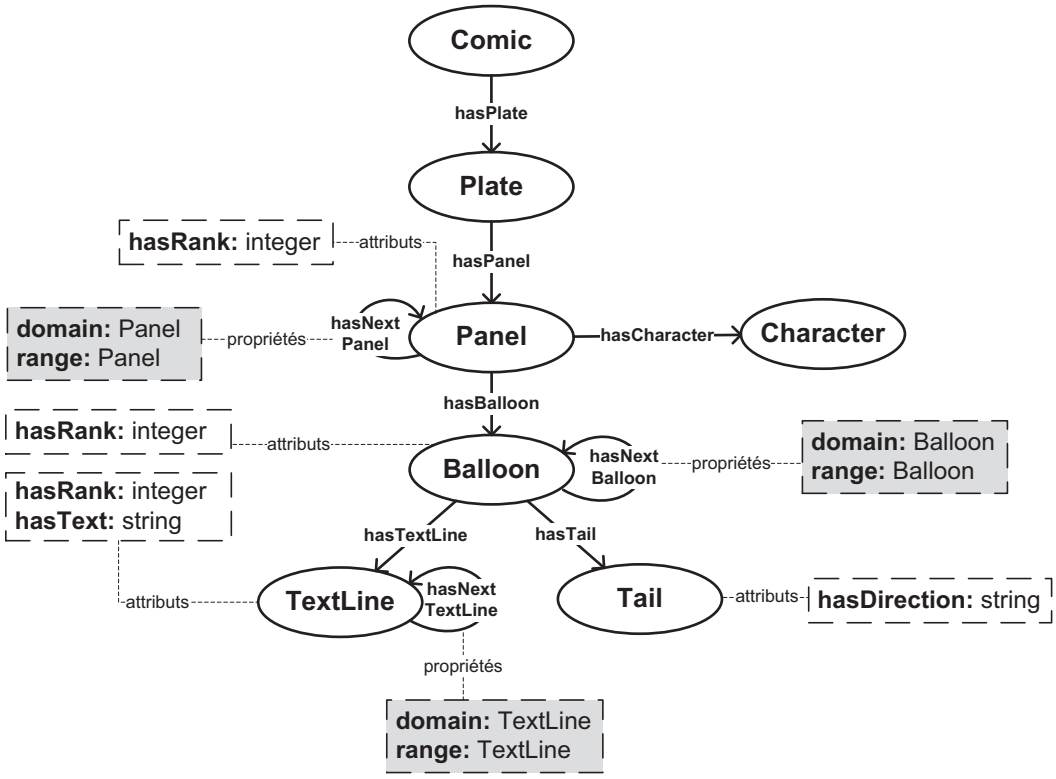


FIGURE 4.6 – Modèle présenté en figure 4.5 enrichi des concepts *Panel*, *Balloon*, *TextLine* et *Character*.

Les propriétés *hasBalloon* et *hasCharacter* sont définies formellement de manière similaire et représentent le lien d'appartenance existant entre, d'une part, une case et, d'autre part, une instance de *Balloon* et de *Character*. La propriété *hasTextLine* représente l'appartenance d'une ligne de texte à un phylactère. Nous avons ici opté pour une conceptualisation très contrainte mettant de côté certains cas de figure non négligeables, notamment celui des lignes de texte hors phylactères. Cela se justifie dans le cadre d'une analyse automatique d'images, comme nous le présentons au chapitre 6.

Une certaine transitivité dans l'inclusion des éléments les uns par rapport aux autres est observable. Les lignes font partie des bulles, faisant elles-mêmes, tout comme les personnages, parties des cases. Les cases sont regroupées au sein de planches qui, mises les unes à la suite des autres, forment un album. L'exploitation de ces appartenances successives trouve un intérêt certain dans un contexte de recherche d'informations au sein d'une base de connaissances. Une requête portant sur les albums où apparaissent dans une même case plusieurs personnages prononçant certains termes pourrait par exemple s'affranchir des notions de planche et de bulle.

Cet aspect de la bande dessinée est formalisé à travers l'introduction de la pro-

priété transitive *hasContent*. Les propriétés *hasPlate*, *hasPanel*, *hasBalloon*, *hasCharacter* et *hasTextLine* sont notées comme étant des sous-propriétés de *hasContent*.

La figure 4.7 illustre les caractéristiques des propriétés présentées.

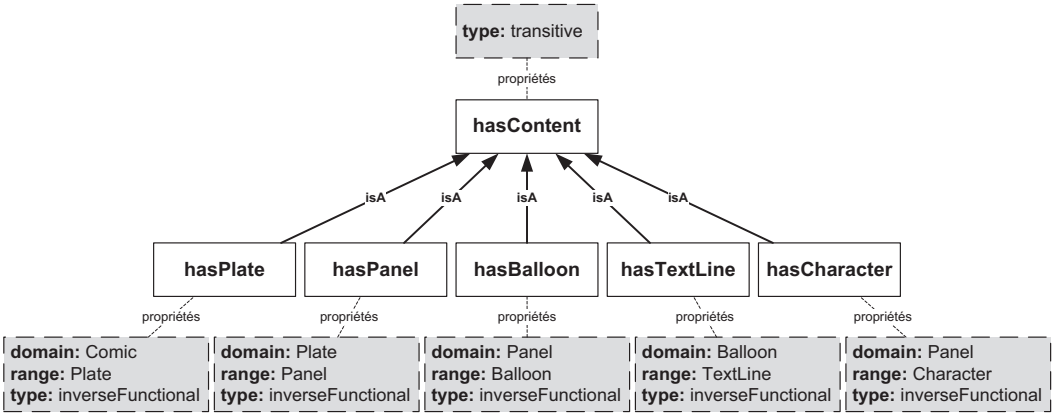


FIGURE 4.7 – Propriétés du modèle de la bande dessinée

4.3.3 Spécialisation du contenu

La sémantique de certains concepts reste d'un degré de granularité assez grossier avec la conceptualisation présentée jusqu'à présent. Les phylactères peuvent notamment se catégoriser en deux sous-ensembles. D'un côté les bulles émises par des personnages (à voix haute ou pensées) ou des éléments de la scène (radio, télévision, etc.), d'un autre côté les cadres de narration. La forme des phylactères varie d'un auteur ou d'une époque à l'autre. Un élément semble faire consensus pour discriminer les bulles des encadrés, c'est la présence, ou non, d'une queue pointant vers la source du son.

Les concepts *NarrativeBalloon* et *SpeechBalloon* sont introduits afin de représenter respectivement les phylactères dont le contenu fait progresser la narration d'une manière ou d'une autre (encadrés, dialogue, etc.), et les bulles matérialisant spécifiquement une section de discours, qu'il soit pensé ou récité à voix haute. Ces éléments sont caractérisés par la présence de lignes de texte, les bulles de discours étant en plus équipées d'une flèche pointant vers la source du son. Ils sont définis par les formules 4.9 et 4.10.

$$\text{NarrativeBalloon} \equiv \text{hasTextLine } \textbf{some } \text{TextLine} \tag{4.9}$$

$$\text{SpeechBalloon} \equiv \text{NarrativeBalloon } \textbf{and } \text{hasTail } \textbf{some } \text{Tail} \tag{4.10}$$

La sémantique des lignes de texte appartenant à ces nouvelles classes de phylactères peut alors également être affinée en conséquence. Certaines lignes sont porteuses d'éléments de discours, tandis que d'autres sont, d'une manière plus générale, au service de la narration. Les concepts introduits pour représenter ces deux notions sont *SpokenTextLine* et *NarrativeTextLine*. Ils sont simplement définis comme des lignes de texte appartenant respectivement à une instance de *SpeechBalloon* ou de *NarrativeBalloon*, tel que le montre la formule 4.11. La propriété *isLineOf* est définie comme étant l'inverse de la propriété *hasLine*.

$$\text{SpokenTextLine} \equiv \text{TextLine } \textbf{and } \text{isLineOf } \textbf{some } \text{SpeechBalloon} \tag{4.11}$$

$$\text{NarrativeTextLine} \equiv \text{TextLine } \textbf{and } \text{isLineOf } \textbf{some } \text{NarrativeBalloon} \tag{4.12}$$

Les bulles de dialogue sont généralement émises par un personnage présent dans la case. Le lien entre un personnage et une bulle de dialogue est exprimé à travers la propriété *says*, ayant pour domain *Character* et pour range *SpeechBalloon*. Le concept *Speaker* représente les personnages émettant des bulles de dialogue selon la formule 4.13.

$$\text{Speaker} \equiv \text{Character } \textbf{and } \text{says } \textbf{some } \text{SpeechBalloon} \tag{4.13}$$

La figure 4.8 illustre les relations de subsomption introduites sur les concepts *Balloon*, *TextLine* et *Character*.

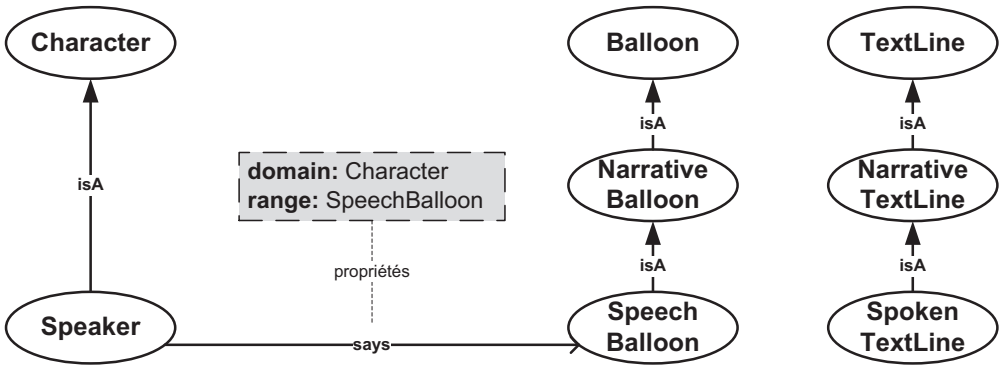


FIGURE 4.8 – Spécialisation des concepts *Character*, *Balloon* et *TextLine*

4.4 Interaction entre les deux ontologies

Nous allons voir dans cette section comment connecter l'ontologie image, notée ici $\mathcal{O}_{image}$, et l'ontologie de la bande dessinée, $\mathcal{O}_{bd}$, des sections 4.2 et 4.3 afin qu'elles puissent communiquer et que leurs capacités de raisonnement respectives puissent être exploitées. Ces deux modèles ont été développés en vue d'être utilisés dans un système d'analyse d'images de bandes dessinées. Dans ce contexte, nous considérons que l'image numérisée d'une planche est la matière première fournie à un algorithme pour qu'il puisse extraire son contenu.

Nous pouvons d'ores et déjà poser une relation d'équivalence entre le concept *Image* de $\mathcal{O}_{image}$ et le concept *Plate* de $\mathcal{O}_{bd}$.

$$\text{Image}_{image} \equiv \text{Plate}_{bd} \tag{4.14}$$

Comme cela a été souligné précédemment, un algorithme d'analyse d'images a pour objectif la détection de caractéristiques visuelles discriminant un type d'objet particulier. Dans notre cas, les extracteurs développés concernent les cases, les bulles, la queue des bulles, le texte et les personnages. Nous pouvons nous servir de ces informations pour créer un lien entre des concepts de $\mathcal{O}_{bd}$ et une formule logique conçue à partir des éléments de $\mathcal{O}_{image}$. On peut donc poser les équivalences :

$$\begin{aligned}
\text{Panel}_{bd} &\equiv \text{ROI}_{image} \textbf{ and } \text{hasExtractor } \textbf{some } (\text{hasROIType } \textbf{value } \text{"Panel"}) \\
\text{Balloon}_{bd} &\equiv \text{ROI}_{image} \textbf{ and } \text{hasExtractor } \textbf{some } (\text{hasROIType } \textbf{value } \text{"Balloon"}) \\
\text{Tail}_{bd} &\equiv \text{ROI}_{image} \textbf{ and } \text{hasExtractor } \textbf{some } (\text{hasROIType } \textbf{value } \text{"Tail"}) \\
\text{TextLine}_{bd} &\equiv \text{ROI}_{image} \textbf{ and } \text{hasExtractor } \textbf{some } (\text{hasROIType } \textbf{value } \text{"TextLine"}) \\
\text{Character}_{bd} &\equiv \text{ROI}_{image} \textbf{ and } \text{hasExtractor } \textbf{some } (\text{hasROIType } \textbf{value } \text{"Character"})
\end{aligned}$$

Les concepts de l'ontologie $\mathcal{O}_{bd}$ représentant le contenu des planches sont alors automatiquement instanciés par les régions d'intérêt issues des extracteurs correspondants. L'attribut $hasROIType_i$ étant fonctionnel, l'unicité du lien entre une instance de ROI_{image} et un concept de $\mathcal{O}_{bd}$ est assurée. Ces derniers sont alors enrichis des propriétés de ROI_{image}, permettant ainsi de manipuler les éléments d'une bande dessinée dans le repère de la planche.

La figure 4.9 illustre ces interactions entre les ontologies.

4.5 Conclusion

Deux ontologies ont été présentées dans ce chapitre, la première formalisant les notions mises en œuvre lors de l'étape d'analyse d'images, l'autre formalisant le

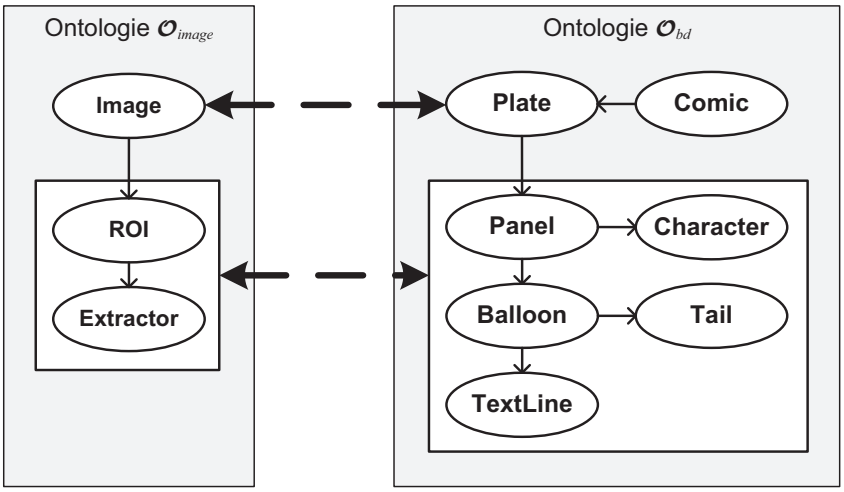


FIGURE 4.9 – Interaction entre les ontologies image et bande dessinée

domaine de la bande dessinée.

L'ontologie image possède l'avantage d'être totalement indépendante d'un quelconque contexte d'application. Elle permet de manipuler les données d'entrée à analyser, d'organiser les extractions issues de différents algorithmes et d'en évaluer les performances par rapport à des données de référence. Le format des concepts représentant les régions d'intérêt et des relations spatiales est compatible avec les standards existants.

L'ontologie BD est composée de concepts permettant de rendre compte de la structure classique d'une bande dessinée. Cette conceptualisation a été développée dans un objectif d'analyse d'images et s'interface avec l'ontologie image à travers des relations d'équivalence entre certains de leurs concepts. Nous verrons dans la suite de ce document que certains points peuvent être amendés afin d'adapter tel ou tel aspect du modèle à différents cadres d'application.

Chapitre 5

Construction d'une vérité terrain

Sommaire

5.1 Introduction 85
5.2 Description du contenu 86
5.2.1 Les albums 86
5.2.2 Les fichiers images 86
5.2.3 Les pages 87
5.2.4 Les cases 87
5.2.5 Les phylactères 87
5.2.6 Le texte 88
5.3 Annotations visuelles 88
5.3.1 Les cases 88
5.3.2 Les bulles 89
5.3.3 Les lignes de texte 89
5.3.4 Les personnages 91
5.4 Annotations sémantiques 91
5.4.1 Les pages 91
5.4.2 Les cases 92
5.4.3 Les bulles 92
5.4.4 Les lignes de texte 93
5.5 Stockage et évaluation 93
5.5.1 Structure de la vérité terrain 93
5.5.2 Évaluation de l'homogénéité 97
5.6 Conclusion 98

5.1 Introduction

Afin d'évaluer la justesse des travaux menés conjointement dans le cadre du projet eBDtheque, nous avons recherché dans la littérature l'existence d'une base de bande dessinées annotées pouvant servir de référence. Nous avions besoin d'une segmentation des composants principaux de chaque page (cases, bulles, texte) et d'une indication sur leur ordre dans la séquence de lecture afin d'être en mesure d'évaluer à la fois les travaux menés en traitement d'images et ceux menés en analyse sémantique.

L'état de l'art s'est trouvé vierge de tout contenu sur ce sujet et nous avons été contraints à constituer notre propre base d'expérimentations et à l'annoter afin d'en faire une vérité terrain à part entière. La construction d'une vérité terrain est toujours coûteuse en terme de temps de travail humain. C'est particulièrement vrai lorsqu'elle implique la segmentation minutieuse d'un grand nombre de zones dans des images. Les cent pages de notre base ont été traitées en une journée par une vingtaine de volontaires travaillant au laboratoire L3i. Afin de disposer d'une base commune pour l'évaluation des travaux portant sur l'analyse de la bande dessinée, cette vérité terrain a été rendue disponible pour la communauté scientifique via le site du projet eBDtheque [1] [Guérin 2013]. Son contenu, ainsi que le protocole de construction, aussi bien sur les aspects de segmentation d'images que sur les aspects d'annotation sémantique, sont détaillés dans ce chapitre.

5.2 Description du contenu

Comme dit en introduction, Scott McCloud a défini la bande dessinée comme étant des "images picturales et autres fixes volontairement juxtaposées en séquences, destinées à transmettre des informations et/ou à provoquer une réaction esthétique chez le lecteur" [McCloud 1993]. Cette définition est volontairement assez large pour englober dans son spectre la majorité des œuvres produites jusqu'à aujourd'hui. La composition de la base doit refléter cette hétérogénéité de forme pour donner à tout un chacun l'opportunité de confronter ses algorithmes à un panel globalement représentatif du visage de la bande dessinée mondiale. Nous avons donc contacté des auteurs aux styles visuels différents afin d'obtenir des droits d'exploitation sur un ensemble d'œuvres diverses et variées. Nous avons ensuite sélectionné un corpus de cent pages dont les caractéristiques sont détaillées ci-après.

5.2.1 Les albums

Les albums desquels sont tirées les pages composant la base ont été publiés entre 1905 et 2012. 29 pages du corpus font parties d'album édités avant 1953 et 71 d'albums sortis dans les années 2000. La qualité du papier, la saturation des couleurs et les textures des aplats liées aux évolutions des techniques d'impression peuvent grandement varier d'une page à l'autre. Les œuvres sont majoritairement d'origine française (81%), puis américaine (13%) et enfin japonaise (6%). Les styles vont de la bande dessinée classique franco-belge au manga japonais en passant par le webcomics et la bande dessinée classique américaine.

5.2.2 Les fichiers images

Les fichiers images composant la base ont en partie été fournis par la société A3DNum qui a été mandatée pour la numérisation de 14 albums. Sur l'ensemble des fichiers, numérisés à une résolution de 600 points par pouce et encodés au format

1. http ://ebdtheque.univ-lr.fr

PNG avec une compression sans pertes, nous avons retenu 46 pages pour intégrer le corpus. Les 54 images restantes ont été sélectionnées depuis onze sites internet d'auteurs de webcomics aux styles différents. Celles-ci sont encodées au format JPEG, compressées avec pertes, et leur résolution varie de 72 à 300 points par pouce. Cette hétérogénéité de la qualité du corpus d'images permet de tester la robustesse des algorithmes aux dégradations de la qualité des supports.

5.2.3 Les pages

Les pages présentent elles-mêmes des caractéristiques très hétérogènes. Parmi l'ensemble, 72 sont imprimées en couleurs avec, selon les auteurs et les époques, une majorité d'aplats, des aquarelles et des zones colorées à la main. Sur les 28 restantes, 16 présentent une coloration en niveaux de gris et 12 sont simplement en noir et blanc. L'un des albums retenu possédant deux versions pour chaque page, l'une en couleur, l'autre en noir et blanc, nous avons intégré à la base des exemples de chacune d'entre elles afin de pouvoir comparer les performances d'un algorithme sur un même style graphique en éliminant la composante couleur. Cinq des 100 fichiers images contiennent une double page. Nous avons donc, strictement parlant, 105 et non 100 pages dans notre base, chacune d'entre elles présentant une structure distincte.

5.2.4 Les cases

Les cases contenues dans les pages sont également de formes variées. Bien que la plupart d'entre elles soient délimitées par une ligne noire, une proportion non négligeable présente une configuration ou au moins une partie de la case est indistinguable du fond de la page. Deux pages sont composées uniquement de cases sans bordure, la délimitation visuelle se faisant grâce au contraste entre le fond de la case et le fond de la page. Neuf pages contiennent des cases se chevauchant, douze ne contiennent que des cases sans bordure et un grand nombre possède des cases reliées par un objet à cheval sur plusieurs d'entre elles.

5.2.5 Les phylactères

Les phylactères de la base contiennent aussi leur lot de diversité. Certains d'entre eux sont complètement fermés, d'autres partiellement. Ils sont de fond blanc ou jaune et peuvent prendre la forme d'une bulle ovale, rectangulaire, avoir une bordure lisse, en forme de « nuage » ou en « pointes ». La plupart d'entre eux sont dotés d'une queue pointant vers le locuteur, mais certains en sont dépourvus. Du texte est présent en dehors de toute bulle sur 33 pages du corpus et 8 d'entre elles ne contiennent tout simplement aucune bulle.

5.2.6 Le texte

Le texte présent sur les pages est écrit à l'ordinateur sur 61% d'entre elles, le reste étant écrit à la main. La plupart des pages sont issues d'œuvres françaises, où le texte est écrit en français. Seules 13 pages contiennent exclusivement du texte en langue anglaise et 6 pages en kanjis japonais. Des onomatopées sont visibles dans 18 pages.

Au total, les 100 pages qui composent cette base de données représentent un ensemble de 849 cases, 1091 phylactères et 4667 lignes de texte.

5.3 Annotations visuelles

Afin de couvrir un champ d'applications le plus large possible, nous avons choisi de détourer les cases, les bulles, les lignes de texte et les personnages. Chacun de ces éléments a été délimité par une boîte englobante rectangulaire dont les quatre côtés sont parallèles aux bords de l'image. Nous avons choisi ce degré de granularité afin de limiter au maximum la part de suggestivité de la personne effectuant l'annotation. La bande dessinée étant un art extrêmement hétérogène dans sa forme et notre base intégrant volontairement des échantillons d'albums pouvant être qualifiés de non-conventionnels, laisser une trop grande liberté d'interprétation sur la forme, parfois suggérée, des éléments conduirait à une variance des annotations pénalisant l'homogénéité de la vérité terrain. Cette granularité de segmentation n'est pas marginale et se retrouve par ailleurs dans plusieurs bases largement utilisées, telles que VOC [Everingham 2009] et [Yao 2007].

Les phylactères ont, dans un second temps, également été détourés le long de leur contour, afin d'intégrer une première mesure pour les travaux portant sur la détection des queues pointant vers les locuteurs. Nous décrivons ci-après la définition que nous faisons des différents éléments à extraire et le protocole à suivre pour leur extraction.

5.3.1 Les cases

Nous appelons cases les zones de l'image, généralement de forme rectangulaire, représentant une scène. Il y a toujours au minimum une case par page, la page entière pouvant faire office de case le cas échéant. Lorsqu'une case est délimitée par une bordure noire, la boîte englobante se place au plus près possible de son cadre. Les pages n'ayant pas toujours été numérisées parfaitement droites, il est parfois impossible d'avoir une boîte englobante collant exactement à la bordure. Lorsque le cadre de la case est partiellement absent ou suggéré par les cases voisines, la boîte englobante vient circonscrire le contenu de la case. Dans tous les cas, les éléments (bulles, texte, dessins) sortant du cadre sont tronqués (voir figure 5.1).

FIGURE 5.1 – Exemple d'annotation de cases : ce strip contient trois cases dont celle du milieu suggérée par les deux autres. La segmentation est effectuée sans tenir compte des éléments sortants du cadre.

5.3.2 Les bulles

Nous définissons une bulle comme une zone de l'image englobant une ou plusieurs lignes de texte, graphiquement délimitée par une frontière physique identifiable ou suggérée par la présence d'une flèche pointant vers le locuteur. Bien que rare, les bulles vides, c'est à dire ne contenant aucune ligne, sont tolérées et également segmentées si elles sont clairement identifiables par leur contour ou par une flèche. La segmentation au pixel suit le contour de la bulle (voir figure 5.2c) tandis que la segmentation par boîte englobante ne considère que le corps du phylactère et tronque la flèche. En effet, celle-ci peut parfois traverser toute la case, générant ainsi une boîte englobante peu représentative de la position de l'élément recherché (voir figure 5.2a). Lorsque la bulle n'est pas fermée, c'est à dire lorsque son contour reste partiellement ouvert, la segmentation se fait au plus proche du contenu (voir figure 5.2b).

5.3.3 Les lignes de texte

Les lignes de texte sont définies comme une suite de caractères alignés dans une même orientation (voir figure 5.3a). Cette définition englobe aussi bien les textes de dialogue ou de narration, très souvent situés à l'intérieur de bulles, que les onomatopées directement incrustées dans l'image de la case. Le texte de ces dernières, bien que n'étant qu'occasionnellement parallèle aux bords de la case, est tout de même détouré par une boîte englobante horizontale par souci d'homogénéité à travers la base. La bande dessinée étant un art graphique statique, l'expression des émotions d'un personnage se fait par l'action conjointe du dessin et du texte, parfois sous la forme d'un unique caractère. Un point d'exclamation pour la surprise ou un point d'interrogation pour l'incompréhension par exemple. Ces caractères isolés sont porteurs d'information et sont donc segmentés comme des lignes de texte à part entière (voir figure 5.3b). De même, nous avons fait le choix d'intégrer dans cette catégorie les textes illustratifs, tels que le « STOP » d'un panneau de signalisation routière ou la devanture d'un magasin (voir figure 5.3c). Bien qu'à la frontière entre le texte constitutif du récit et la représentation graphique de la scène, ces éléments sont tout

(a) (b) (c)

FIGURE 5.2 – Différents exemples d'annotation de bulles : (a) par boîte englobante n'intégrant pas la flèche, (b) par boîte englobante sur une bulle non fermée, (c) au pixel près.

de même invariablement lus par le lecteur et leur détection est donc potentiellement intéressante pour de multiples usages, notamment l'analyse de scène.

(a) (b) (c)

FIGURE 5.3 – Différents exemples d'annotation de texte : (a) une ligne de texte classique, (b) un caractère unique, (c) du texte illustratif non horizontal.

5.3.4 Les personnages

Les personnages sont le dernier type d'éléments à avoir été annotés pour la base eBDtheque. La notion de *personnage* peut donner cours à un grand nombre d'interprétations différentes lorsqu'elle est associée à la bande dessinée et doit donc être précisée. Les personnages d'une BD ne sont pas nécessairement des êtres humains, ou humanoïdes, ni même des êtres vivants. Quand bien même, il serait peu pertinent d'annoter chaque être humain apparaissant dans une case alors que certains ne sont rien d'autre qu'une partie du décor. Nous avons donc choisi de limiter l'annotation aux personnages principaux d'une bande dessinée, c'est à dire ceux émettant au moins une bulle de dialogue au cours de l'album. Leur boîte englobante a été définie de manière à maximiser l'espace occupé par le personnage à l'intérieur de la boîte. Par conséquent, certaines parties des personnages, comme les jambes ou les bras, ne sont pas, dans certains cas, à l'intérieur de la boîte (voir figure 5.4).

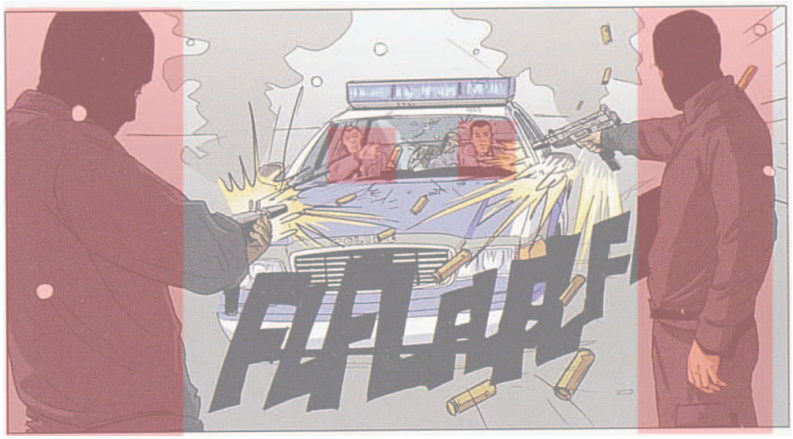


FIGURE 5.4 – Exemple d'annotation des personnages : les bras des tireurs ne sont pas inclus dans les boîtes englobantes afin de maximiser l'espace occupé par le personnage.

5.4 Annotations sémantiques

Une fois les éléments segmentés, chacun d'entre eux a été enrichi d'informations précisant la sémantique de leurs caractéristiques visuelles.

5.4.1 Les pages

Les pages ont été annotées selon trois dimensions. La première de ces dimensions est celle de l'objet page en tant qu'individu composant une bande dessinée. Les informations s'y rapportant sont celles venant spécifier la manière dont elles doivent être parcourues. La direction de lecture, de gauche à droite ou de droite à gauche pour les œuvres japonaises, est précisée à travers la métadonnée `readingDirection`. La

langue principale du texte présent dans la page est également entrée, ainsi que le fait que l'image représente une double page ou non, respectivement par les métadonnées `language` et `doublePage`.

Deuxièmement, des annotations au caractère plus techniques, telles que la résolution (`resolution`) de numérisation et la définition (`height` et `width`), sont venues enrichir les objets pages en tant qu'images numérisées, en tant que contenu informatique. Ces annotations correspondent respectivement aux attributs du concept *Image* : *hasResolution*, *hasHeight* et *hasWidth*. Enfin, afin de préciser le contexte de chaque page de la base, ces dernières ont également été annotées d'informations bibliographiques relatives aux albums desquels elles ont été tirées. Ce contexte est utile pour toutes les études transversales, portant par exemple sur l'évolution de la taille des cases chez un éditeur à travers le temps, ou encore sur les différences de verbosité entre deux auteurs aux styles distincts. Le nom de la collection (`collectionTitle`), de l'album (`albumTitle`) auquel appartient la page, ainsi que sa date de publication (`releaseDate`) ont été saisis. De même, le nom du scénariste (`writerName`), du dessinateur (`drawerName`) et de l'éditeur (`editorName`) sont également renseignés. Afin de faciliter la tâche de celui n'ayant pas accès aux images mais souhaitant utiliser la base, nous fournissons également, lorsqu'ils sont disponibles, l'ISBN (`isbn`) et le site web (`website`) permettant d'identifier l'album.

5.4.2 Les cases

Les cases sont tout d'abord classées selon leur position dans la séquence de lecture de la page. Une métadonnée `rank` leur a été attribuée, prenant pour valeur un entier fixé entre 1 et n, le premier de la séquence prenant la valeur 1 et le dernier la valeur n pour une page contenant n cases.

5.4.3 Les bulles

Les bulles ont également été classées selon leur ordre de lecture dans la page. Pour une page contenant m bulles, la métadonnée `rank` de chacune d'entre elles a été fixée à une valeur entre 1 et m. Les bulles n'ont pas été classées relativement aux cases les contenant car les relations d'appartenance des bulles aux cases n'ont pas été renseignées dans cette première version de la base.

Une deuxième métadonnée a été introduite afin de qualifier le type de trait composant le contour de la bulle. Cette caractéristique est en effet souvent porteuse d'information, notamment sur la manière dont le texte que contient la bulle est prononcé par son locuteur. Nous avons choisi de nous limiter à trois possibilités couvrant la majorité des cas de figure (voir figure 5.5) : *smooth* pour les bulles de dialogue au contour lisse, *wavy* pour les bulles de pensées au contour nuageux et *zigzag* pour les bulles au contour hérissé synonyme d'émotion ou exprimant un cri. L'attribut `borderStyle` de chaque bulle prendra donc sa valeur parmi l'une de ces trois possibilités.

Deux autres informations ont été entrées pour chaque bulle, elles concernent

la flèche pointant le personnage source du phylactère. La première porte sur la position de son extrémité et est stockée dans la métadonnée `tailTip`. La seconde concerne l'orientation du dernier segment de la flèche. Les valeurs possibles pour cette métadonnée, baptisée `tailDirection`, sont les huit relations cardinales *N, NE, E, SE, S, SO, O, NO*, plus une neuvième valeur indiquant l'absence de flèche, *none*. De plus, lorsque le personnage émetteur est visuellement identifiable, le lien entre la bulle et celui-ci est renseigné.

FIGURE 5.5 – Les différents types de contour de bulle, de gauche à droite et de haut en bas : wavy, zigzag, smooth, other.

5.4.4 Les lignes de texte

Les lignes de texte ont tout d'abord été retranscrites textuellement en lettres capitales. Dans un second temps, la fonction de chaque ligne dans la narration a été spécifiée à travers la métadonnée `textType`. Nous avons recensé six catégories distinctes. D'un côté il y a le texte qui sert à raconter l'histoire et qui peut être soit parlé (`speech`), soit pensé (`thought`) ou soit narré (`narrative`). De l'autre côté il y a le texte qui apporte de l'information ponctuelle et contextuelle. Il s'agit des onomatopées (`onomatopoeia`) et du texte diégétique (`illustrative`), c'est à dire le texte faisant partie de l'univers de la scène représentée (plaque d'immatriculation, devanture de magasin, etc.). La sixième catégorie correspond aux notes (`note`) incrustées dans la page, telles que la signature de l'auteur, le numéro de la planche, le titre du strip, etc. Tout ce qui n'est, en définitive, ni parlé, ni pensé, ni narré, ni une partie du dessin, ni une onomatopée.

Enfin, lorsque une ligne de texte fait partie d'un phylactère, une référence à ce dernier est ajouté en métadonnée.

5.5 Stockage et évaluation

5.5.1 Structure de la vérité terrain

Afin de faciliter la diffusion et l'utilisation de la base auprès de la communauté scientifique, nous avons choisi de stocker ces annotations dans des fichiers au format

SVG (Scalable Vector Graphics). Ce format, ouvert et bien documenté, présente, outre le fait d'être activement maintenu par un groupe de travail du W3C, deux avantages majeurs pour la diffusion d'une base telle que la nôtre.

Premièrement, il fournit un moyen extrêmement simple de visualiser les éléments segmentés. Nul besoin d'installer de logiciels tels que Matlab, Adobe Illustrator ou leurs équivalents « open sources » pour afficher leur contenu. La lecture du format SVG qui est avant tout, comme son nom l'indique, un format de dessin vectoriel, est en effet supportée par tous les navigateurs internet récents. Basé sur le format XML, il est compatible avec les feuilles de style en cascade (CSS), ce qui permet d'enrichir considérablement la clarté de visualisation de la base pour un utilisateur en définissant, par exemple, des styles différents pour chaque type d'éléments (voir figure 5.6). Il est ainsi possible d'appréhender visuellement la base très rapidement, en n'affichant au besoin que les couches pertinentes pour un usage donné.

(a) (b) (c)

FIGURE 5.6 – Le style d'affichage des éléments annotés de la base peut être paramétré. Par exemple en rouge pour les cases (a), en bleu pour les bulles (b) et en vert pour le texte (c).

Deuxièmement, la syntaxe à base de balises héritée de XML rend l'intégration des annotations sémantiques au sein des éléments très simple, notamment grâce à la balise standard `metadata`.

A chaque image de notre base est donc associé un unique fichier SVG contenant l'intégralité de ses annotations. La granularité la plus basse avec laquelle nous traitons est donc la page, le concept d'album ne se retrouvant qu'à travers les métadonnées, les pages d'un même album ne sont regroupées d'aucune façon. A la suite de l'entête XML standard contenant notamment les informations d'encodage du fichier, l'élément racine est une balise `<svg>` ayant pour fils une balise `<title>`,

contenant l'identifiant du document, et cinq balises `<svg>`. Ces dernières contiennent les annotations collectées sur les cinq types d'éléments concernés, la page, les cases, les bulles, le texte et les personnages. Le type d'élément contenu par une balise est spécifié par son attribut `class`. La première balise, `class="Page"`, contient les descriptions relatives à la page et possède pour cela deux balises filles. La première, `image`, possède plusieurs attributs dont un spécifiant le lien vers le fichier image de la page (`xlink:href`) et deux précisant la largeur et la hauteur de cette image (`width` et `height`). La deuxième, `metadata`, contient les informations bibliographiques relatives à l'album et les propriétés de la page décrites en section 5.4.

Les quatre balises suivantes contiennent respectivement les annotations relatives aux cases, aux bulles, aux lignes de texte et aux personnages. Leur structure est identique quel que soit le type d'élément concerné. Chacun d'entre eux est spécifié par une balise `<polygon>` qui permet de définir un élément graphique dans l'image dont les coordonnées sont renseignées au sein de l'attribut `points`. C'est la valeur de cet attribut qui, une fois convertie au format WKT, viendra qualifier la propriété *hasWKT* du concept *ROI* de notre modèle (voir sous-section 4.2.1). Chaque élément `<polygon>` possède un attribut `id` utilisé comme référence pour les liens entre bulle/personnage et texte/bulle. Ces éléments `<polygon>` possèdent également chacun une balise fille `metadata` contenant en attributs les informations relatives à la classe à laquelle appartient l'élément, telles que décrites en section 5.4.

La référence de la bulle contenant une ligne est attachée à l'attribut `idBalloon`, ajouté à la balise `metadata` de l'élément `Line` correspondant. Un personnage pouvant émettre plusieurs bulles dans une même case, et une bulle pouvant être émise par plusieurs personnages, les liens entre bulles et personnages sont spécifiés d'une manière différente. Une nouvelle balise `<svg>`, ayant pour classe `linkBC`, est introduite à la suite des éléments contenant les personnages. Celle-ci a pour fils un ensemble de `<polygon>` matérialisant chacune par une ligne la liaison entre une bulle et un personnage. Une référence à la bulle et au personnage correspondant sont enregistrées pour chaque balise sous la forme des attributs `idBalloon` et `idCharacter`.

L'encadré 5.1 présente un exemple du contenu de l'un de ces fichiers.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<svg>
  <title>CYB_BUBBLEGOM_T01_005</title>
  <svg class="Page">
    <image x="0" y="0" width="750" height="1060" href="CYB_BUBBLEGOM_T01_005.jpg">
    <metadata collectionTitle="collection" albumTitle="titre"
              drawerName="dessinateur" writerName="scenariste"
              editorName="editeur" ISBN="979-10-90655-01-0"
              website="http://website.com" releaseDate="2009"
              language="french" readingDirection="leftToRight"
              doublePage="false" pageNumber="5" resolution="300"/>
  </svg>
  <svg class="Panel">
    <polygon points="53,95 268,95 268,292 53,292 53,95"
             id="CYB_BUBBLEGOM_T01_005_PANEL_0">
      <metadata rank="1"/>
    </polygon>
    ...
  </svg>
  <svg class="Balloon">
    <polygon points="61,103 137,105, 143,103, 150,141, 143,172 61,172 61,103"
             id="CYB_BUBBLEGOM_T01_005_BALLOON_0">
      <metadata borderStyle="smooth"
                boundingBox="61,103 143,103 143,172 61,172 61,103"
                tailTip="494,214"
                tailDirection="SE"
                rank="1"/>
    </polygon>
    ...
  </svg>
  <svg class="Line">
    <polygon points="373,121 432,121 432,132 373,132 373,121"
             id="CYB_BUBBLEGOM_T01_005_LINE_0">
      <metadata textType="speech"
                idBalloon="CYB_BUBBLEGOM_T01_005_BALLOON_0">
        COMME TOI.
      </metadata>
    </polygon>
    ...
  </svg>
  <svg class="Character">
    <polygon points="84,153 261,153 261,298 84,298 84,153"
             id="CYB_BUBBLEGOM_T01_005_CHARACTER_0"/>
    ...
  </svg>
  <svg class="linkBC">
    <polygon points="143,150 144,150 136,253 135,253 143,150"
             idCharacter="CYB_BUBBLEGOM_T01_005_CHARACTER_0"
             idBalloon="CYB_BUBBLEGOM_T01_005_BALLOON_0"/>
  </svg>
</svg>
```

ENCADRE 5.1 – Exemple de fichier SVG constituant la base eBDtheque

5.5.2 Évaluation de l'homogénéité

Puisqu'une vingtaine de personnes ont pris part, à différents degrés d'implication, à la création de cette base, la précision de la segmentation des éléments n'est probablement pas parfaitement homogène. En effet, celle-ci a pu varier d'une personne à l'autre, chacun ayant sa propre sensibilité aux données et chacun ayant intégré les consignes de segmentation d'une manière plus ou moins nette. Nous avons donc cherché à mesurer cette homogénéité sur l'ensemble de la base en demandant à chacun en fin de session de traiter les cases d'une même page, sélectionnée au préalable pour ses caractéristiques représentatives de la base (voir figure 5.8). Elle contient dix cases dont quatre délimitées par une bordure, une partiellement close et une dernière sans aucune bordure.

Nous avons commencé par définir ce que serait une marge d'erreur acceptable pour la position d'un coin d'une case. Les pages étant de définitions différentes, un seuil t, relatif aux dimensions de la page, de 0,5% a été fixé, ce qui correspond pour notre image de test, d'une définition de 750 par 1060 pixels, à un écart maximal de plus ou moins cinq pixels en ordonnée et de plus ou moins quatre pixels en abscisse. Nous avons ensuite redonné aux vingt personnes les instructions fournies lors de la construction de la base, c'est-à-dire un détourage au plus près du contenu de la case, excluant les bulles et tronquant les éléments dépassant hors du cadre. Cette manipulation a produit vingt ensembles de quarante points (quatre points par case) desquels nous avons tiré un ensemble moyen, contenant la position moyenne en x et en y de chacun des quarante points. La distance de chaque point à sa position moyenne a alors été calculée. Le nombre de points compris à une certaine distance de cette moyenne est reporté sur le graphique représenté en figure 5.7.

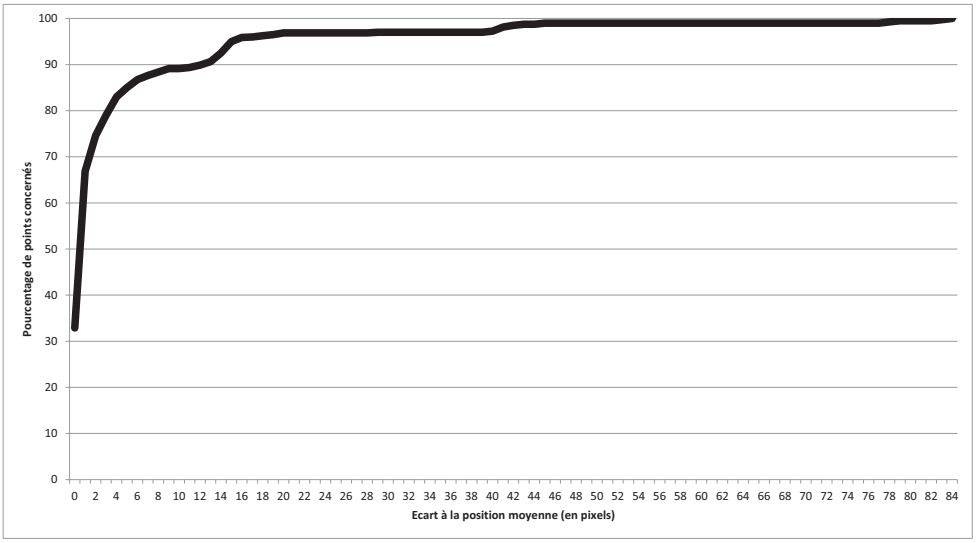


FIGURE 5.7 – Ce graphique représente le nombre de points inclus à une certaine distance de leur position moyenne en ordonnée. La courbe en abscisse est similaire.

Si l'on considère le seuil t de 0,5%, on constate qu'environ 85% des points sont dans la marge d'erreur acceptable que nous avons définie et peuvent donc être considérés comme homogènes à travers le groupe de personnes. L'écart-type moyen des quarante points est de 0,15% (1,13 pixels) en largeur et de 0,12% (1,28 pixels) en hauteur. Les deux sursauts visibles sur la courbe aux abscisses 15 et 40 sont dus à la mauvaise segmentation de treize cases, incluant ou excluant des zones qui n'auraient pas dû l'être. La figure 5.8 illustre en hachuré la différence entre les zones labellisées par tout le monde et les zones labellisées par au moins une personne. Les erreurs dues à une mauvaise compréhension des consignes sont cependant facilement et rapidement identifiables et ont pu être corrigées lors d'une phase de post-traitement de la base.

Bien que l'erreur n'ait été mesurée que sur la segmentation des cases, les résultats sont raisonnablement extensibles aux boîtes englobantes des bulles, le protocole d'extraction étant très similaire. La segmentation des bulles au niveau pixel et la segmentation des personnages ayant été assurée par une seule personne, son homogénéité n'est soumise qu'à la régularité de la personne au cours du temps et est, par conséquent, difficilement mesurable.

5.6 Conclusion

Le développement de la base eBDtheque a été l'objet d'un travail continu sur trois années. Elle décrit de manière visuelle et sémantique les cases, bulles, lignes de texte et personnages d'une centaine de pages de bandes dessinées au contenu varié. Les différentes annotations ont été intégrées progressivement, en fonction des besoins de l'équipe de recherche ainsi que des retours de la communauté. Cela s'est traduit par la mise en ligne de plusieurs versions consécutives, la première ayant fait l'objet d'une présentation complète dans un article publié à l'occasion de l'édition 2013 de la conférence ICDAR. Nous recevons régulièrement des requêtes de téléchargement et nous savons que la base est actuellement utilisée internationalement par des équipes situées en Chine, au Liban, en Inde et en Allemagne.

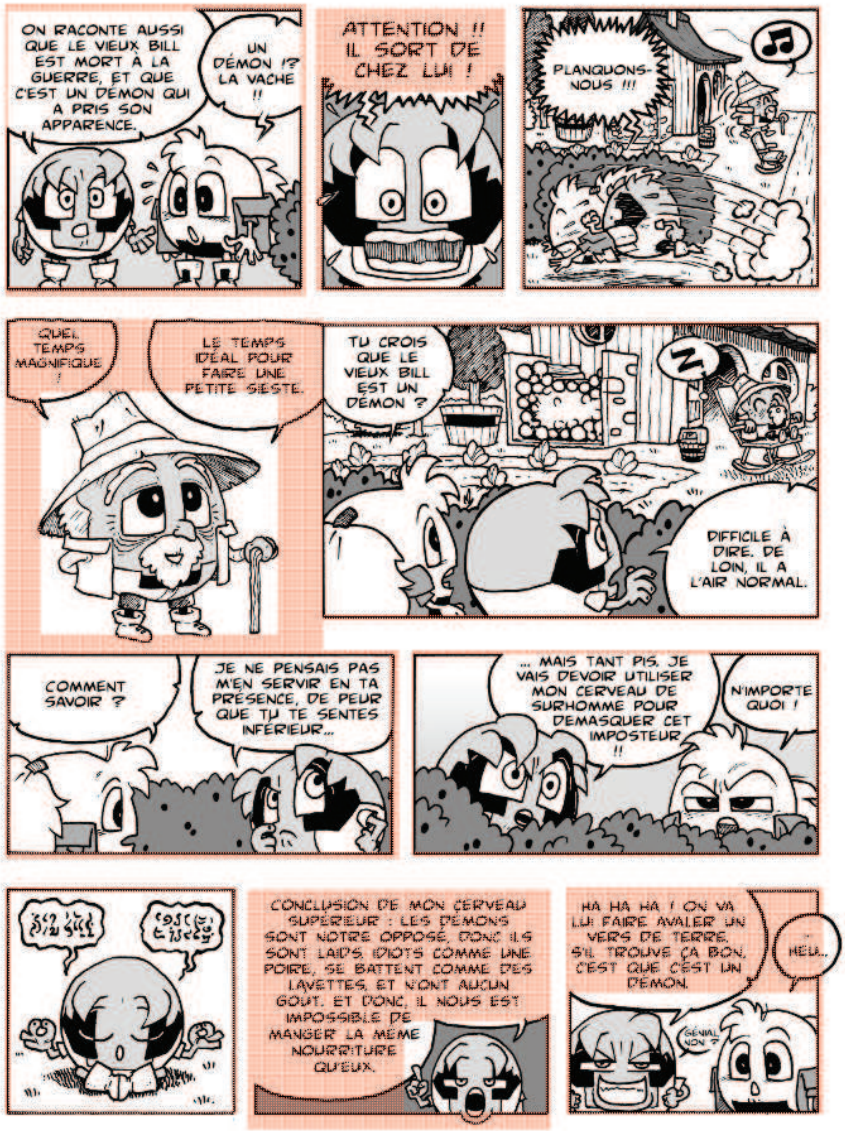


FIGURE 5.8 – Les zones hachurées représentent la différence entre les zones labellisées comme étant une case pour tous les participants et les zones labellisées par au moins un participant.

Troisième partie

Vers une compréhension globale du contenu d'une bande dessinée.

Chapitre 6

Analyse contextuelle des planches

Sommaire

6.1 Introduction 103
6.2 Description générale 104
6.2.1 Communication entre les deux systèmes 104
6.2.2 Contraintes pour l'extraction 105
6.3 Boucle d'interaction 107
6.3.1 Reconnaissance des éléments simples 108
6.3.2 Reconnaissance des éléments complexes 111
6.4 Évaluation des performances du système 117
6.4.1 Protocole d'évaluation 117
6.4.2 Évaluations individuelles 119
6.4.3 Évaluation générale 121
6.5 Conclusion 122

6.1 Introduction

Nous présentons dans ce chapitre le point de jonction entre cette thèse et celle de Christophe Rigaud [Rigaud 2014a], menée en parallèle au sein du projet eBDtheque. Le cadre présenté dans la deuxième partie de ce manuscrit est mis en œuvre pour l'analyse et la compréhension progressive du contenu d'une page de bande dessinée. Une telle compréhension du contenu est cruciale lorsque l'on se place dans un contexte d'automatisation du processus d'adaptation de bandes dessinées imprimées vers de nouveaux supports de lecture. Nous proposons un framework complet basé sur une dynamique de bouclage et de communication entre le système de traitement d'images et le cadre formel. Il est fait référence à la partie algorithmique de traitement d'images sous le nom de *système bas niveau*, ou *SBN*, sous entendu proche des données pixellaires. Le système formé des ontologies développées, couplées à un moteur d'inférence, est, quant à lui, dénommé ci-après *système haut niveau*, ou *SHN* pour plus de clarté.

6.2 Description générale

6.2.1 Communication entre les deux systèmes

Nous considérons ici les ontologies $\mathcal{O}_{image}$ et $\mathcal{O}_{bd}$ développées comme parties intégrantes d'un *système haut niveau* fournissant un cadre d'interprétation au contenu extrait par un système de traitement d'images bas niveau. Le but du *système haut niveau* est d'interagir de manière itérative avec le *système bas niveau* pour comprendre progressivement le contenu d'une page, de ses éléments les plus simples aux éléments plus complexes. L'ensemble du *SHN*, représenté par le schéma de la figure 6.1, inclus donc les deux ontologies présentées dans le chapitre 4, formant ensemble notre base de connaissances. Cette base, une fois peuplée par des données en provenance du *système bas niveau*, est traitée par un moteur d'inférence pour en extraire des conclusions logiques, dépendantes des données et de leur cohérence par rapport à la connaissance formalisée.

Ces conclusions peuvent consister en la validation d'éléments extraits, en leur rejet ou la création d'individus renvoyés au système d'extraction d'images.

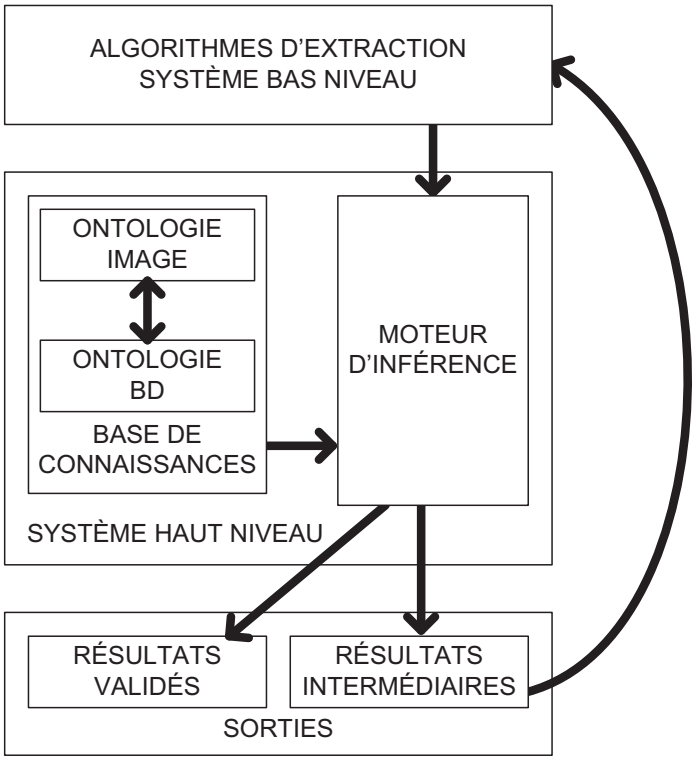


FIGURE 6.1 – Représentation des interactions entre les systèmes bas et haut niveau

Les algorithmes bas niveau ont été développés pour extraire des éléments spécifiques depuis l'image entière d'une planche ou une région de celle-ci. Les systèmes

haut et bas niveau interagissent afin que la totalité des extractions et leurs relations soient en concordance avec la représentation de la connaissance du domaine étudié. La figure 6.2 illustre la boucle de traitement liant les deux systèmes et un exemple concret d'application est présenté en section 6.3.

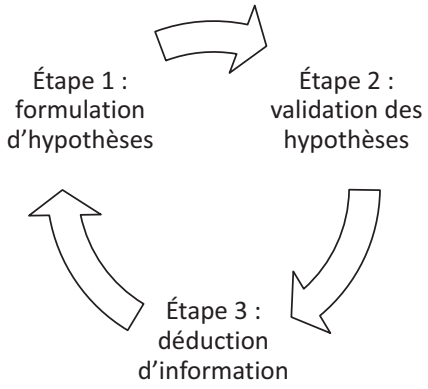


FIGURE 6.2 – Boucle de traitement

A l'étape 1, le *système bas niveau* propose au *système haut niveau* des hypothèses tenant en un premier ensemble de régions segmentées dans l'image. Ces régions sont labellisées selon leur type supposé (cases, bulles, etc.). Lors de la deuxième étape, le *SHN* évalue ces hypothèses, valide celles qu'il juge correctes et supprime celles qu'il considère erronées. A la troisième étape, de nouvelles informations sont déduites à partir des hypothèses validées, mises en perspectives de la connaissance du domaine. Celles-ci sont renvoyées au *SBN* qui s'en sert, au cours d'une deuxième itération, pour extraire des éléments plus complexes, tels que les personnages.

6.2.2 Contraintes pour l'extraction

L'ontologie que nous avons développée nous permet de catégoriser les éléments d'une bande dessinée selon quatre types : les cases, les bulles, le texte et les personnages. Ce ne sont pas les seuls composants d'une BD mais cela représente tout de même une large part du contenu des planches et une tâche d'analyse d'images conséquente. Comme nous l'avons vu au chapitre 1, la bande dessinée est un art au sens pur du terme. Elle possède ses codes, ses traditions, mais ne suit aucune spécification formelle transversale à absolument toutes les œuvres créées. Certaines bandes dessinées ne contiennent pas de cases, d'autres ne contiennent pas de bulles ni de texte. Une conceptualisation valide pour toutes les œuvres, notamment concernant les liens entre les éléments du contenu, serait trop générale pour être d'une quelconque utilité pratique dans le cadre de l'analyse automatique de ce contenu.

La conceptualisation que nous avons adoptée met, pour rappel, en exergue les points suivants :

- une case *appartient* à une et une seule planche,
- une bulle *appartient* à une et une seule case,
- une occurence de personnage *appartient* à une et une seule case,
- une ligne de texte *appartient* à une et une seule bulle.

Ces contraintes représentent les conventions établies entre les auteurs et les lecteurs dans un souci de clarté et de lisibilité. Elles sont valides pour une grande majorité de planches de bande dessinée, à quelques exceptions près. La place des éléments doit en effet être clairement identifiable au premier coup d'œil, ce qui implique, par exemple, que les bulles et les personnages soient entièrement inclus dans les cases. S'il n'est cependant pas rare d'observer des débordements à l'extérieur du cadre des cases, ceux-ci sont, la plupart du temps, réduits au minimum. Ainsi, même si la relation d'appartenance spatiale entre les éléments peut parfois être sujette à caution, le lien sémantique reste non ambigu.

Le terme *appartient* symbolise donc la situation où un élément graphique en chevauche un autre (a fortiori lorsqu'il est inclus dans un autre) sur une proportion *significative* de sa surface. En cas d'intersections ou d'inclusions multiples, seule celle avec la plus petite région est conservée. Une ligne de texte doit par exemple être considérée comme appartenant à la bulle qui la contient, avant d'appartenir à la case contenant cette bulle.

Face au manque de données concernant les proportions d'intersection des éléments dans la bande dessinée, nous avons estimé ce que serait cette proportion *significative* en utilisant notre jeu de données annotées. Pour chaque élément de la base eBDtheque, nous avons observé quel serait le type de son conteneur direct, pour une proportion donnée de sa surface devant être recouverte afin d'établir une relation d'appartenance.La figure 6.3 illustre, pour chaque type d'éléments (cases, bulles, texte et personnages), le nombre d'individus possédant un conteneur valide selon nos contraintes, en fonction de cette proportion de recouvrement. La procédure a été appliquée séparément sur chaque type d'éléments car il n'y a pas de raison de supposer que le ratio de recouvrement optimal soit transversal à la nature des éléments. Les quatre ratios optimaux correspondent donc, pour notre base, aux pics respectifs des courbes de chaque type d'éléments.

En sélectionnant les valeurs de recouvrement donnant les résultats optimaux pour chaque type d'éléments, nous obtenons un modèle représentatif de 99,6% des cases, 87,4% des bulles, 81,6% des lignes de texte et 94,9% des personnages de la base eBDtheque.

Le taux plus faible des bulles s'explique par le choix, effectué lors de la création de la vérité terrain, de segmenter les cases non fermées au plus proche de leur contenu et en ne tenant pas compte des bulles (cf section 5.3.1). Ces dernières se retrouvent alors en dehors de toute case et sont attribuées à la planche.

De même, les 81% de conformité des lignes de texte est due à la présence non

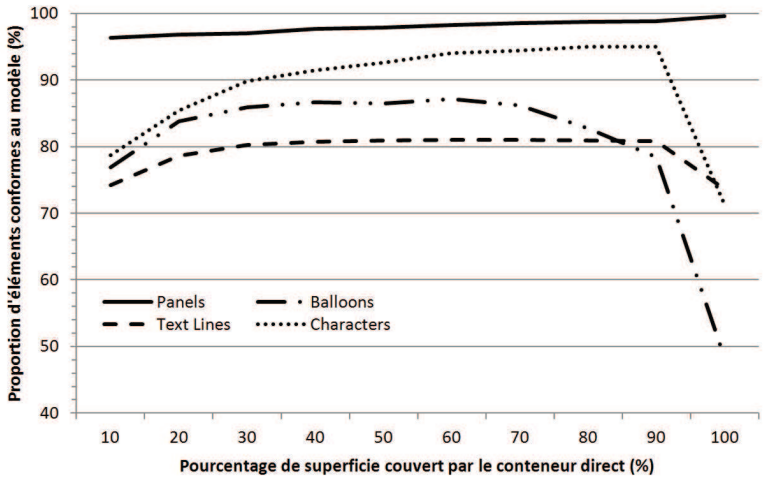


FIGURE 6.3 – Pourcentage de cases, bulles, lignes de texte et personnages de la base eBDtheque correspondant à notre conceptualisation, en fonction de leur surface couverte par leur conteneur direct.

négligeable dans le jeu de données de webcomics ayant tendance à ajouter des informations bibliographiques au sein de chaque planche. Ces cas particuliers gonflent le nombre de lignes textuelles laissées de côté, s'ajoutant par ailleurs aux onomatopées et au texte illustratif.

La chute observable pour les bulles, le texte et les personnages à une valeur de recouvrement égale à 100% correspond à un effet de bord bien compréhensible. Les éléments devant être entièrement inclus dans leur conteneur, le moindre pixel en dehors de celui-ci entraine l'attribution de la relation d'appartenance à l'élément supérieur. C'est en revanche une bonne chose pour les cases, celles-ci devant appartenir à la planche qui est, par définition, l'élément racine et n'est incluse dans aucun autre.

6.3 Boucle d'interaction

Le processus global est constitué de deux itérations de la boucle de traitement présentée en figure 6.2. Lors de la première itération sont traités les éléments dits « simples », c'est-à-dire les cases, les phylactères et les lignes de texte. Ceux-ci sont qualifiés de simples car il possèdent une structure relativement régulière et ont déjà bénéficié de l'attention d'un certain nombre de travaux. Lors de la deuxième itération de la boucle, la détection des personnages, plus complexes de par leur diversité visuelle, est abordée.

Les avancées progressives effectuées à chaque étape du processus sont illustrées à travers le couple de cases présenté en figure 6.4.

FIGURE 6.4 – Exemple suivi illustrant la compréhension progressive du contenu d'une planche. Crédits : Jay Scott Pike - Boy Love Girls 41 - Lev Gleason.

6.3.1 Reconnaissance des éléments simples

6.3.1.1 Première itération, étape 1 (hypothèses)

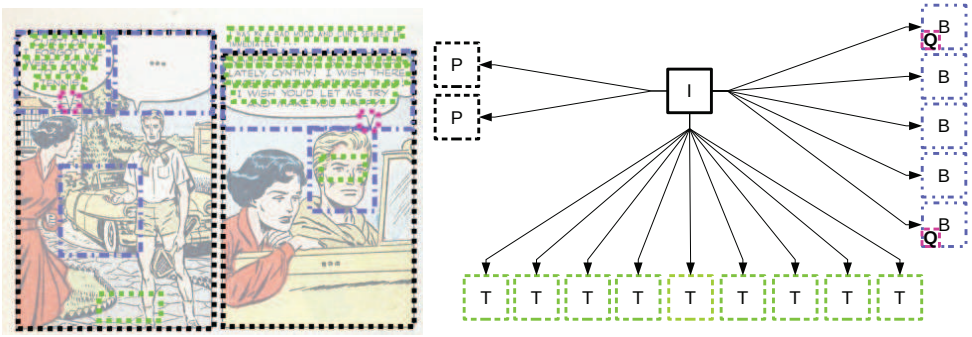
Les algorithmes d'extraction sont lancés sur l'image afin d'en extraire le contenu. Le fonctionnement de ces algorithmes ne sera pas décrit en détails dans ce manuscrit. Un aperçu des différentes approches mises en œuvre est disponible en section 3.3. Le lecteur intéressé par le sujet pourra se référer à la thèse de Christophe Rigaud [Rigaud 2014a], ainsi qu'aux articles cités dans la section d'évaluation 6.4.2, pour une description complète de chaque traitement.

Lors de cette première itération, l'attention est portée sur les cases, les bulles, les queues des bulles et les lignes de texte. Tous ces éléments forment les hypothèses de départ devant être validées par le *système haut niveau*. Celles-ci sont introduites dans la A-Box via l'assertion d'axiomes instanciant les concepts *ExtractorAuto* et *ROI*. Chaque région d'intérêt est reliée à son extracteur d'origine via la propriété *hasExtractor*. L'attribut *hasROIType* de chaque extracteur est renseigné par le type des éléments pour l'extraction desquels il a été conçu.

Les éléments en pointillés dans la figure 6.5a représentent les hypothèses initiales parmi lesquelles d'éventuelles erreurs de détection peuvent apparaître. Ces éléments sont représentés dans le graphe 6.5b, d'une manière, pour l'instant, non structurée.

6.3.1.2 Première itération, étape 2 (validation)

Lors de la deuxième étape, les hypothèses proposées par le *système bas niveau* sont confrontées aux contraintes formalisées dans l'ontologie. Les régions segmentées sont premièrement catégorisées en cases, bulles et lignes de texte grâce aux règles, présentées en section 4.4, interprétées par un moteur d'inférence. Ces axiomes sont insérés dans la A-Box sous la forme de faits. En effet, des incohérences peuvent potentiellement être introduites à l'étape suivante, bloquant ainsi le moteur d'inférence et empêchant de fait la déduction de l'équivalence entre une région d'intérêt et sa



(a) (b)

FIGURE 6.5 – Hypothèses initiales (en pointillés) des composants d'une image I après extraction des cases P, des lignes de texte T, des bulles B et de leur queue Q.

classe dans $\mathcal{O}_{bd}$ à la volée.

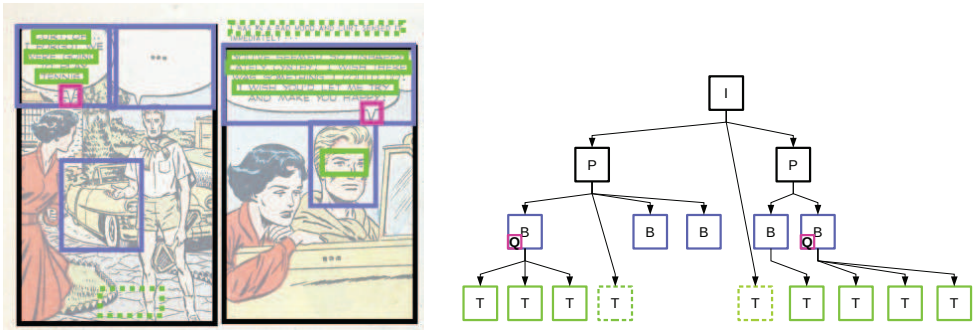
Dans un second temps, chaque élément e est lié à son élément conteneur direct E. Celui-ci est sélectionné parmi l'ensemble des éléments extraits selon la méthode décrite en section 6.2.2. S'il est impossible de trouver un conteneur pour un élément e parmi l'ensemble des régions extraites, la planche est alors considérée comme étant ce conteneur. Des axiomes du type hasContent(?E, ?e) sont introduits dans la A-Box, où la propriété *hasContent* est remplacée par *hasPanel*, *hasBalloon* et *hasTextLine* en fonction du concept instancié par e dans $\mathcal{O}_{bd}$. Afin de vérifier la validité des hypothèses, les éléments E sont simplement pris comme des instances du concept ROI_i. Nous ne tenons pas compte de leur classe dans l'ontologie $\mathcal{O}_{bd}$. Cela a pour effet d'introduire des incohérences entre le type d'un conteneur et le domaine d'une propriété issue de *hasContent*, rendant ainsi le modèle inconsistant. Ces incohérences permettent de mettre en exergue de possibles erreurs de segmentation effectuées lors de l'étape de traitement bas niveau. Un moteur d'inférence est de nouveau lancé sur l'ontologie et les raisons de son inconsistance sont gérées les unes après les autres.

Nous nous sommes concentrés dans ces travaux sur l'optimisation de la fiabilité des résultats. Nous avons fait le choix de privilégier la précision des extractions plutôt que leur caractère exhaustif. La détection de mauvaises classifications, telles qu'une case ayant été segmentée par l'extracteur de bulles, et la mise en avant d'éléments ayant pu être oubliés sont tous deux des perspectives à court termes de ces travaux. A l'heure actuelle, notre système est capable de gérer les raisons d'inconsistance suivantes :

- **Une planche (p) contient une bulle (b) ou une ligne de texte (t)** : b ou t est supprimée ;

- **Une case (c1) contient une case (c2) ou une ligne de texte (t)** : c2 ou t est supprimée ;
- **Une bulle (b) contient une case (c)** : si c contient des bulles et si b contient des lignes de texte alors c est supprimée, sinon b est supprimée ;
- **Une bulle (b1) contient une bulle (b2)** : si b1 ne contient pas de lignes alors b1 est supprimée, sinon b2 est supprimée ;
- **Une ligne de texte (t) contient une case (c)** : si c contient des bulles alors t est supprimée, sinon c est supprimée ;
- **Une ligne de texte (t) contient une bulle (b)** : si b contient du texte alors t est supprimée, sinon b est supprimée ;
- **Une ligne de texte (t1) contient une ligne de texte (t2)** : si t1 contient plus d'une ligne alors t1 est supprimée, sinon t2 est supprimée.

Le terme *contient* renvoie dans la liste ci-dessus à la relation d'appartenance d'un élément à un autre, décrite en section 6.2.2. La figure 6.6b illustre l'état actuel de l'organisation des éléments extraits. Ceux-ci ont trouvé leur place au sein d'une structure représentative du contenu de la planche. Les éléments validés sont illustrés en traits continus tandis que les éléments rejetés sont restés en pointillés.



(a) (b)

FIGURE 6.6 – Validation des hypothèses respectant les contraintes de notre conceptualisation. Les éléments validés sont en traits continus.

6.3.1.3 Première itération, étape 3 (déduction)

A cette étape de la première boucle, les incohérences ont été résolues et les éléments extraits restants sont validés et conformes à notre modèle. Ils sont organisés au sein de la hiérarchie de concepts de $\mathcal{O}_{bd}$. L'ontologie étant consistante, un moteur d'inférence est donc en mesure de classer les instances de *Balloon* et de *TextLine*, respectivement en *SpeechBalloon*, *NarrativeBalloon* et *SpokenTextLine*, *NarratedTextLine*, en fonction de leurs caractéristiques.

La classification de ces instances est illustrée sur la figure 6.7.

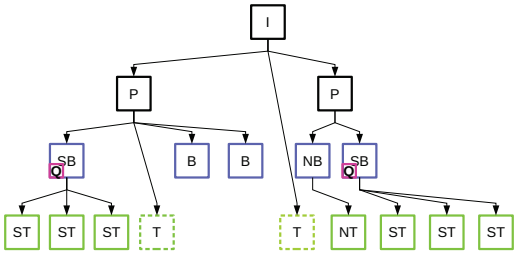


FIGURE 6.7 – Classification des instances en *SpeechBalloon* SB, *NarrativeBalloon* NB, *SpokenTextLine* ST et *NarrativeTextLine* NT à partir de leurs propriétés et des contraintes de $\mathcal{O}_{bd}$.

6.3.2 Reconnaissance des éléments complexes

6.3.2.1 Deuxième itération, étape 1 (hypothèses)

Cette étape marque le début de la deuxième boucle dans laquelle nous nous intéressons aux éléments de l'image plus complexes que sont les personnages. Le *système haut niveau* possède à ce point une certaine connaissance du contenu de la planche, issue de l'analyse effectuée lors de la première boucle. Cette connaissance est utilisée pour formuler des hypothèses sur la localisation probable de personnages dans les cases pour guider leur détection par le *système bas niveau*. Nous nous intéressons dans un premier temps aux personnages principaux de la planche, c'est-à-dire à ceux émettant au moins une bulle de dialogue. Il est raisonnable de considérer qu'une bulle et le personnage l'émettant appartiennent à une seule et même case. Il est donc possible de restreindre l'espace de recherche du personnage à l'intérieur de la zone de l'image représentant cette case. De plus, la position du personnage dans la case peut être estimée par rapport à la direction pointée par la flèche caractéristique des bulles de dialogue.

Nous commençons par extraire de l'ontologie les éléments nécessaires à cette tâche. La requête SPARQL de l'encadré 6.1 permet d'extraire le modèle décrivant les cases contenant des bulles de dialogues.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX img: <http://l3i.univ-larochelle.fr/eBDtheque/image#>
PREFIX bd: <http://l3i.univ-larochelle.fr/eBDtheque/domain#>

DESCRIBE ?plate ?panel ?balloon ?tail
WHERE {
  ?plate rdf:type bd:Plate .
  ?panel rdf:type bd:Panel .
  ?balloon rdf:type bd:SpeechBalloon .
  ?tail rdf:type bd:Tail .
  ?plate bd:hasPanel ?panel .
  ?panel bd:hasBalloon ?balloon .
  ?balloon bd:hasTail ?tail
```

```
}
```

ENCADRE 6.1 – Requête extrayant le modèle contenant toutes les cases d'une page contenant des bulles de dialogue

Les instances de *Panel*, *Balloon* et *Tail* faisant également toutes partie de l'extension du concept *ROI*, chacune possède un attribut *hasWKT* prenant pour valeur les coordonnées de l'élément dans l'image, et donc, dans la planche.

Pour chaque instance de *Tail*, la valeur de l'attribut *hasDirection* est calculée en fonction de la valeur de ses coordonnées. Ces coordonnées sont de type `LINESTRING(x1 y1, x2 y2)`, où `(x1 y1)` et `(x2 y2)` sont les coordonnées respectives, de la base du dernier segment et de l'extrémité de la queue. La détection du dernier segment de la flèche est une caractéristique souhaitée de l'extracteur. En effet, il arrive fréquemment que la queue d'une bulle soit représentée par une ligne brisée. La direction pointée par le dernier segment n'est alors pas nécessairement la même que celle indiquée par le vecteur liant la base à la pointe de la flèche. La figure 6.8 illustre cette situation.

(a) (b)

FIGURE 6.8 – Exemple de flèche complexe, seul le dernier segment de la flèche porte l'information de direction vers l'origine du son. (a) En rouge le vecteur directionnel de la base à la pointe de la queue. (b) En vert le vecteur de la base du dernier segment à la pointe. Crédits : Cyb - Magicien Loose - Studio Cyborga.

L'angle directionnel d'une flèche est reporté dans un espace correspondant à l'échantillonnage du cercle trigonométrique en huit secteurs disjoints. Ceux-ci ont chacun un angle d'ouverture égal à $\pi/4$, correspondant aux valeurs cardinales Nord, Sud, Est, Ouest, Nord-Est, Nord-Ouest, Sud-Est, Sud-Ouest. Une hypothèse sur la position potentielle d'un personnage locuteur est alors émise sous la forme d'une zone d'une certaine taille, incluse dans la case à laquelle appartient la bulle de dialogue, et positionnée dans un espace pointé par la flèche de cette bulle.

Les dimensions de cette zone se doivent d'être bornées d'une certaine façon. Le fait est que la taille de la zone qu'occupe réellement un personnage au sein d'une

case est difficile à estimer a priori. Dans le cas d'une case présentant un gros plan, la quasi-totalité de l'espace peut être occupée par la seule tête d'un personnage. Au contraire, dans le cas d'un plan large, plusieurs personnages peuvent être représentés et n'occuper individuellement qu'une petite partie de la surface de la case.

Nous avons décidé de fixer les dimensions maximales des hypothèses en fonction de la taille moyenne $\overline{s}$ des bulles présentes dans la case selon l'équation 6.1.

$$\overline{s} = \frac{\sum_{i=0}^{n} Wb_i + \sum_{i=0}^{n} Hb_i}{2n} \tag{6.1}$$

Les termes Wb_i et Hb_i représentant respectivement la largeur et la hauteur en pixels de la bulle i, n correspondant au nombre de bulles dans la case courante.

La taille et la position de la région rectangulaire contenant potentiellement un personnage est ensuite définie par les coordonnées de son coin supérieur gauche, $\mathrm{H1}_{x,y}$, et de son coin inférieur droit, $\mathrm{H2}_{x,y}$, selon les formules 6.2.

$$\begin{aligned}
\mathrm{H1}_x &= t_x - O_x(d).\min(\overline{s}, t_x - \mathrm{A1}_x) \\
\mathrm{H1}_y &= t_y - O_y(d).\min(\overline{s}, t_y - \mathrm{A1}_y) \\
\mathrm{H2}_x &= \mathrm{H1}_x + \min(\overline{s}, \mathrm{A2}_x - \mathrm{H1}_x) \\
\mathrm{H2}_y &= \mathrm{H1}_y + \min(\overline{s}, \mathrm{A2}_y - \mathrm{H1}_y)
\end{aligned} \tag{6.2}$$

Où t_x, t_y, ont pour valeur l'abscisse et l'ordonnée de l'extrémité de la flèche. Comme expliqué en section 4.2.1, nous travaillons dans un repère orthonormé dont l'origine correspond au coin supérieur gauche de l'image. Les termes $\mathrm{A1}_{x,y}$ et $\mathrm{A2}_{x,y}$ représentent les coordonnées des coins supérieur gauche et inférieur droit de la case courante. Afin que la zone calculée ne déborde pas du cadre de la case, ses dimensions sont calculées en fonction de la valeur minimale existant entre la dimension $\overline{s}$ et l'écart de t_x, y au bord de la case.

Les fonctions O_x et O_y font correspondre à une valeur d'orientation d un coefficient de décalage en abscisse et en ordonnée permettant de positionner la région d'intérêt correctement vis à vis du phylactère. Les valeurs de ces coefficients sont reportées dans le tableau 6.1.

d	N	NE	E	SE	S	SW	W	NW
O_x	0.5	0.25	0	0.25	0.5	0.75	1	0.75
O_y	1	1	0.5	0	0	0	0.5	1

TABLEAU 6.1 – Coefficient de décalage de la zone en x et en y pour une orientation de la flèche donnée.

La figure 6.9 illustre les huit situations envisagées.

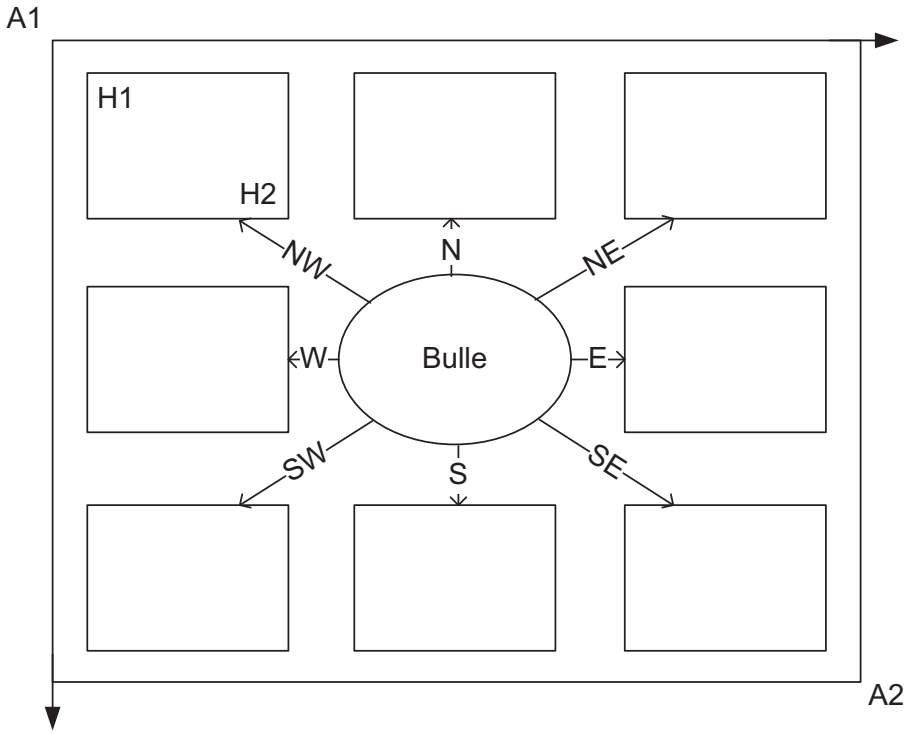


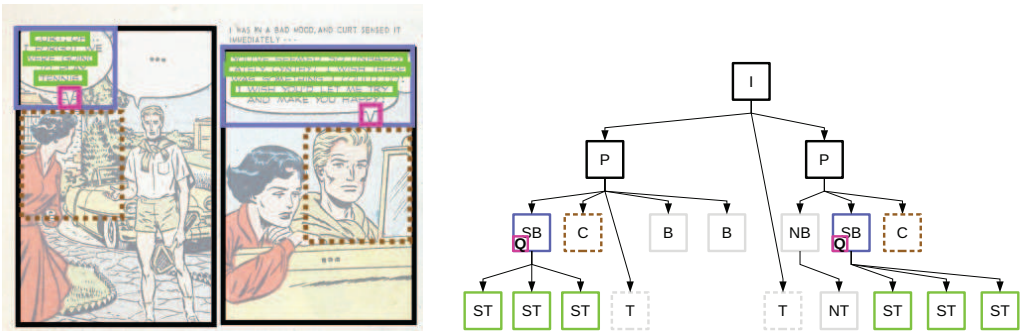
FIGURE 6.9 – Localisation de la zone hypothétique H du locuteur en fonction de la direction de la flèche d'une bulle de dialogue et du cadre de la case A.

Ces hypothèses sont intégrées au modèle (voir figure 6.10) et envoyées au *système bas niveau*. Ces régions servent alors de point de départ pour la recherche des personnages entiers. Les algorithmes mis en œuvre pour la réalisation de cette tâche sont détaillés dans les travaux de thèse de [Rigaud 2014a] et ont été publiés dans [Rigaud 2014c].

Une fois que les régions d'intérêt contenant les personnages ont été calculées par le *système bas niveau*, celles-ci sont retournées au *système haut niveau* pour leur validation (voir figure 6.11).

6.3.2.2 Deuxième itération, étape 2 (validation)

A la manière dont a été évalué le premier ensemble de régions d'intérêt extraites par le système d'analyse bas niveau, ce nouveau lot d'extractions est soumis aux contraintes de notre ontologie. Les régions sont tout d'abord classifiées comme instances du concept $Character_{bd}$ et les axiomes d'instanciation sont assertés dans la A-Box du modèle. Des axiomes *hasCharacter(?E, ?c)* sont ensuite introduits entre chaque personnage et son conteneur direct, tel que décrit à l'étape 2 de la première itération (section 6.3.1.2). La propriété *hasCharacter* ayant pour domaine le concept



(a) (b)

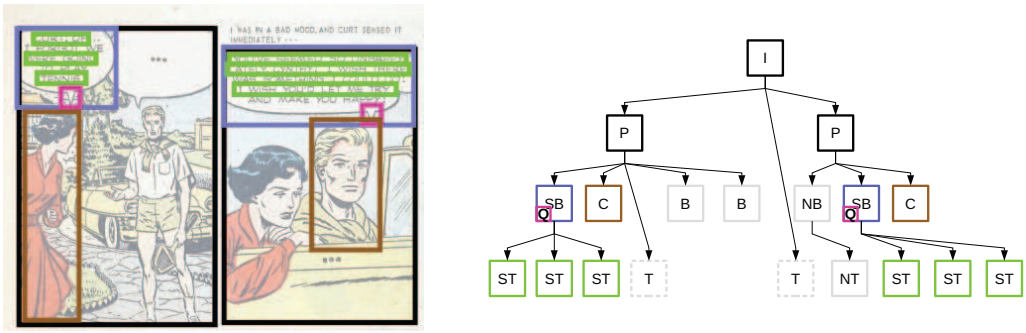
FIGURE 6.10 – Position estimée des régions contenant des personnages C, calculée à partir des bulles de dialogue et des cases. Les régions n'étant pas en relation avec une bulle de dialogue sont masquées dans un souci de lisibilité.

FIGURE 6.11 – Hypothèses affinées par le *système bas niveau* des régions d'intérêt contenant des personnages.

Panel, les éventuelles incohérences insérées sont détectées par le moteur d'inférence et sont traitées une à une.

- **Une planche (p) contient un personnage (i)** : i est supprimé ;
- **Une bulle (b) contient un personnage (i)** : i est supprimé ;
- **Une ligne de texte (t) contient un personnage (i)** : i est supprimé ;
- **Un personnage (i) contient une case (c), une bulle (b) ou une ligne de texte (t)** : i est supprimé ;
- **Un personnage (i1) contient un personnage (i2)** : i2 est supprimé.

Les éléments validés sont représentés en ligne continue sur l'image et le graphique de la figure 6.12.



(a) (b)

FIGURE 6.12 – Validation des personnages C par rapport aux régions fournies par le *système bas niveau* et aux contraintes du *système haut niveau*.

6.3.2.3 Deuxième itération, étape 3 (déduction)

La dernière étape de ce processus consiste à déduire de l'ensemble des personnages validés ceux qui sont effectivement reliés, d'un point de vue scénaristique, à une bulle de dialogue. Nous extrayons pour cela du modèle l'ensemble des cases contenant des bulles de dialogue et des personnages selon la requête 6.2.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX img: <http://l3i.univ-larochelle.fr/eBDtheque/image#>
PREFIX bd: <http://l3i.univ-larochelle.fr/eBDtheque/domain#>

DESCRIBE ?plate ?panel ?balloon ?tail ?character
WHERE {
  ?plate rdf:type bd:Plate .
  ?panel rdf:type bd:Panel .
  ?balloon rdf:type bd:SpeechBalloon .
  ?tail rdf:type bd:Tail .
  ?character rdf:type bd:Character .
  ?plate bd:hasPanel ?panel .
  ?panel bd:hasCharacter ?character .
  ?panel bd:hasBalloon ?balloon .
  ?balloon bd:hasTail ?tail
}
```

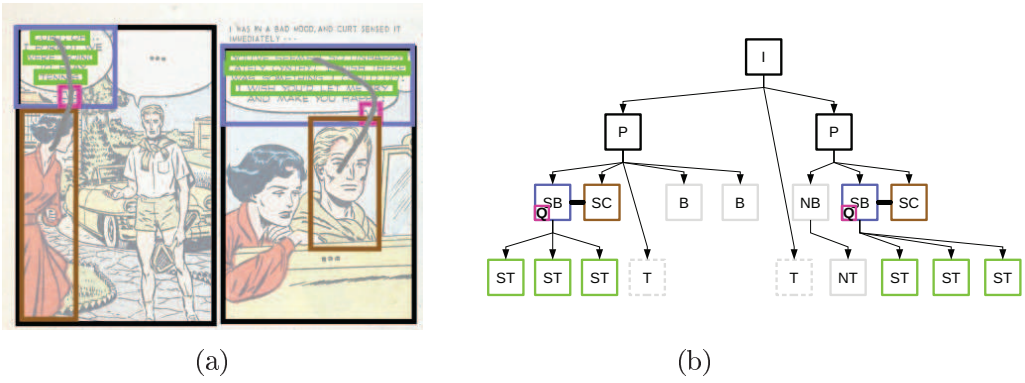
ENCADRE 6.2 – Requête extrayant le modèle comprenant toutes les cases d'une page contenant à la fois des bulles de dialogue et des personnages

Pour chaque case du modèle extrait, nous traitons chacune de ses bulles de dialogue *sb* séquentiellement. A partir des attributs *hasDirection* et *hasWKT*, nous récupérons, pour chaque individu objet de la propriété *hasTail*, la direction d pointée par la flèche ainsi que les coordonnées $t_{x,y}$ de son extrémité. Une ligne abstraite est alors tracée depuis t dans la direction d jusqu'à entrer en intersection avec une région d'intérêt c, instance de *Character*, donnée par l'attribut *hasWKT* hérité du

concept ROI_{image}.

Un axiome *says(c, sb)* est alors asserté dans la base de connaissances, permettant de ce fait au moteur d'inférence de déduire l'instanciation du concept *Speaker* par c, via la règle 4.13.

La figure 6.13 illustre l'ajout de ces informations dans la structure de la page.



(a) (b)

FIGURE 6.13 – Sélection des personnages émetteurs d'une bulle de dialogue et classification dans le concept *Speaker* SC. Les liens entres bulles et personnages sont représentés par un trait gris dans l'image et par un lien plus épais sur le graphe.

6.4 Évaluation des performances du système

Nous présentons dans cette section une évaluation des performances de notre système d'analyse de bandes dessinées. Après avoir détaillé le mécanisme mis en place pour l'évaluation des résultats, nous commenterons les performances du système pour des tâches indépendantes, puis, dans un second temps, sur une chaîne de traitement globale et continue.

6.4.1 Protocole d'évaluation

L'évaluation des méthodes d'extraction et de validation des éléments d'une planche a été menée sur la base eBDtheque présentée dans le chapitre 5. Celle-ci est, pour rappel, composée de cent planches au contenu annoté visuellement et sémantiquement. La composition de la base a volontairement été voulu variée, du point de vue de l'origine des planches (franco-belge, manga, webcomics, etc.), mais également de celui du style de dessin et de mise en page.

Nous avons présenté dans la section 6.2.2 les capacités de notre conceptualisation à représenter formellement le contenu de cette base. Nous avons établi que notre modèle est conforme et apte à représenter 99,6% des cases, 87,4% des bulles, 81,6%

des lignes de texte et 94,9% des personnages. Les caractéristiques des éléments restants rendant ceux-ci incompatibles avec notre modélisation. Ces valeurs font donc office d'objectif absolu dont notre système doit se rapprocher sans pouvoir espérer le dépasser.

Les éléments en provenance de la vérité terrain et des extracteurs automatiques ont été introduits dans la A-Box et classés dans la hiérarchie de concepts de l'ontologie de bande dessinée $\mathcal{O}_{bd}$. Pour chaque intersection entre un individu extrait automatiquement i_t de type t et un individu de la vérité terrain i'_t du même type t, une instance e du concept *Evaluation* est créée et liée à ceux-ci par les propriétés *hasCandidate(e,i_t)* et *hasReference(e,i'_t)*. Chacune de ces instances est enrichie d'un attribut *hasVOCScore* représentant une métrique, introduite par [Everingham 2009]. Celle-ci donne un indice de qualité q basée sur l'intersection et l'union des ensembles comparés selon l'équation 6.3, où ri_t et ri'_t représentent respectivement les ensembles de pixels correspondant aux régions d'intérêt i_t et i'_t.

$$q_{ii'} = \frac{|ri_t \cap ri'_t|}{|ri_t \cup ri'_t|} \tag{6.3}$$

Cette métrique a valeur de jugement, une région i étant considérée validée par rapport à une référence i' si la valeur de $q_{ii'}$ est supérieure à un seuil fixé à $0,5$. La règle 4.3 est modifiée en conséquence en 6.4 afin de classer les différentes évaluations en instances de *Validation* ou *Error*.

$$\text{Validation} \equiv \text{Evaluation } \textbf{and } \text{hasVOCScore } \textbf{some } \text{double}[>=0{,}5] \tag{6.4}$$

$$\text{Error} \equiv \text{Evaluation } \textbf{and } \text{hasVOCScore } \textbf{some } \text{double}[<0{,}5] \tag{6.5}$$

La qualité de détection des éléments est donnée de manière globale via des mesures de rappel R, précision P et de F-Score F, calculées à partir des formules 6.6.

$$R = \frac{\text{nbElementsValides}}{\text{nbElementsVT}} \qquad P = \frac{\text{nbElementsValides}}{\text{nbElementsAuto}} \qquad F = \frac{2RP}{R+P} \tag{6.6}$$

Les valeurs de « nbElementsValides », « nbElementsVT » et « nbElementsAuto » sont obtenues à partir des requêtes SPARQL présentées dans l'encadré 6.3, où `<TYPE>` est remplacé par le nom de la classe des éléments que l'on souhaite évaluer.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bd: <http://l3i.univ-larochelle.fr/eBDtheque/domain#>
PREFIX img: <http://l3i.univ-larochelle.fr/eBDtheque/image#>

SELECT (count(DISTINCT ?element) as ?nbElementsValides)
WHERE {
  ?element rdf:type bd:<TYPE>.
```

```
  ?validation rdf:type img:Validation.
  ?validation img:hasReference ?element
}

SELECT (count(DISTINCT ?element) as ?nbElementsAuto)
WHERE {
  ?element rdf:type bd:<TYPE>.
  ?extractor rdf:type img:AutoExtractor.
  ?element img:hasExtractor ?extractor.
}

SELECT (count(DISTINCT ?element) as ?nbElementsVT)
WHERE {
  ?element rdf:type bd:<TYPE>.
  ?extractor rdf:type img:GroundTruthExtractor.
  ?element img:hasExtractor ?extractor.
}
```

ENCADRE 6.3 – Requête permettant d'obtenir le nombre d'éléments de la vérité terrain et d'un type donné ayant validé une extraction

6.4.2 Évaluations individuelles

Nous avons, dans un premier temps, cherché à évaluer les performances de notre système pour chaque type d'éléments indépendamment des autres. Il faut pour cela que la segmentation des éléments non évalués soit parfaite, de sorte que celle-ci ne puisse pas être la source de potentielles erreurs et interférer avec les résultats d'extraction des éléments évalués. L'ensemble d'évaluation $\mathbb{E}$ est donc défini à partir des ensembles d'extraction automatique $\mathbb{E}_{\text{a}}$ et de vérité terrain $\mathbb{E}_{\text{vt}}$ selon les formules 6.7.

$$\mathbb{E}_{\text{a}} = \{\mathbb{C}, \mathbb{B}, \mathbb{T}, \mathbb{P}\}_{\text{a}} \tag{6.7}$$

$$\mathbb{E}_{\text{vt}} = \{\mathbb{C}, \mathbb{B}, \mathbb{T}, \mathbb{P}\}_{\text{vt}} \tag{6.8}$$

$$\mathbb{E} = \{\mathbb{X}_{\text{a}} \cup \mathbb{E}_{\text{vt}} \setminus \mathbb{X}_{\text{vt}}\} \tag{6.9}$$

Avec $\mathbb{C}$, $\mathbb{B}$, $\mathbb{T}$, $\mathbb{P}$ respectivement des ensembles de cases, bulles, lignes de texte et personnages et $\mathbb{X}$ l'ensemble à évaluer, correspondant à l'un des quatre précédents.

Le tableau 6.2 présente une synthèse des résultats obtenus pour chaque ensemble d'éléments. Ces résultats sont mis en perspective de ceux obtenus sans l'utilisation du *système haut niveau* et donc de la boucle d'interaction présentée dans ce chapitre.

L'apport du *système haut niveau* n'est pas particulièrement notable dans le cas de la détection des cases, principalement car la détection initiale est d'avance correcte. L'écart restant à combler pour une détection parfaite s'explique par la sensibilité des algorithmes aux liaisons entre les cases ainsi qu'à l'absence de gouttière ou de bordure dans le cas de grandes cases.

	R (%)	P (%)	F (%)
Évaluation des cases ($\mathbb{X} = \mathbb{C}$)			
[Rigaud 2014a]	81,24	86,55	83,81
[Rigaud 2014a] + SHN	80,69 **(-0,55)**	87.03 **(+0,48)**	83,74 **(-0,07)**
Évaluation des bulles ($\mathbb{X} = \mathbb{B}$)			
[Rigaud 2014b]	57,90	73,84	64,91
[Rigaud 2014b] + SHN	54,79 **(-3,11)**	88,76 **(+14,92)**	67,75 **(+2,84)**
Évaluation des lignes de texte ($\mathbb{X} = \mathbb{T}$)			
[Rigaud 2013c]	61,00	19,66	29,75
[Rigaud 2013c] + OCR	60,13	42,43	49,75
[Rigaud 2013c] + SHN	44,54 **(-16,46)**	65,05 **(+22,62)**	52,88 **(+3,13)**
Évaluation des personnages ($\mathbb{X} = \mathbb{P}$)			
Ité. 2.1 – Hypothèses	19,28	26,91	22,47
Ité. 2.1 – Extraction	23,31 **(+4,03)**	33,55 **(+6,64)**	27,51 **(+5,04)**
Ité. 2.2 – Validation	21,57 **(-1,74)**	40,52 **(+6,97)**	28,16 **(+0,65)**

TABLEAU 6.2 – Mesures de rappel R, précision P et F-score F pour la reconnaissance des cases, bulles, lignes de texte et personnages. Chaque évaluation met en rapport les résultats obtenus avec et sans l'utilisation du *SHN*.

Un gain de précision notable apporté par le *SHN* est en revanche observable sur la détection des bulles. Malgré une petite baisse dans la valeur de rappel, le F-score est légèrement augmenté. Le rappel moyen est dû à l'utilisation d'un algorithme d'extraction qui, contrairement à celui utilisé dans [Rigaud 2013b], est uniquement capable de détecter les bulles strictement fermées par un contour (voir figure 3.11) mais est en revanche indépendant d'une quelconque détection préalable du texte.

La détection des lignes de texte profite également de manière importante de l'apport du *système haut niveau*. Le gain de précision est plus de 20% supérieur à celui apporté par l'utilisation d'un OCR et de plus de 45% par rapport à la segmentation seule. La perte de rappel, plus importante que dans le cas des bulles, est liée à la faiblesse identifiée de notre conceptualisation rejetant les lignes n'étant pas à l'intérieur de phylactères. Le texte illustratif à l'intérieur des cases, ou les notes détectées dans la page, bien que potentiellement valides sont rejetées au moment de la validation.

Enfin, les personnages bénéficient également d'un gain sensible de précision. Les efforts ayant été concentrés sur la détection des personnages émettant une bulle de dialogue, une grande partie des occurrences d'apparition, correspondant à des personnages sans bulle, a été laissée de coté pour le moment. Sur l'ensemble des personnages de la base eBDtheque, seulement 56,8% d'entre eux sont liés à une

bulle, ce qui explique le rappel plus faible constaté.

La capacité du modèle à affilier un personnage à une bulle de dialogue a également été évaluée. La couverture des personnages de la base eBDtheque atteint 70,66%.

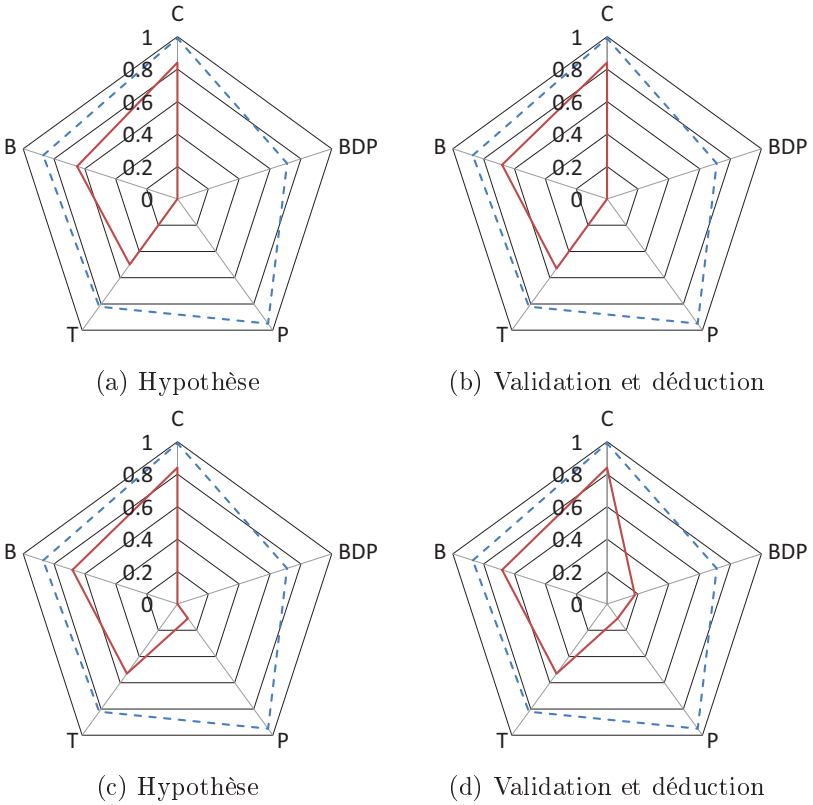
6.4.3 Évaluation générale

Dans un second temps, nous avons évalué la performance de notre système pour une analyse complète des planches. Lors des évaluations individuelles présentées en section 6.4.2, nous nous servions des données de la vérité terrain comme de gardes fous afin d'observer successivement le comportement de notre framework sur chaque type d'éléments. Pour cette évaluation générale, les cases, les bulles, le texte et les personnages sont évalués simultanément et tous les éléments sont issus des algorithmes d'extraction. Aucun élément n'est importé depuis la vérité terrain. La validation de chaque type d'élément est alors étroitement liée à la qualité de l'extraction et de la validation des autres types, notamment de ses conteneurs directs et indirects. La validation de l'extraction d'une ligne de texte est par exemple dépendante de la validation de la bulle la contenant, elle-même conditionnée par la présence d'une case valide. Les erreurs d'extraction et de validation des éléments se propagent et se combinent, impactant en conséquence les résultats obtenus en bout de chaîne, notamment pour les personnages et leurs liens avec les bulles de dialogue.

La figure 6.14 illustre la progression de la quantité d'information découverte sur les planches. La ligne en pointillés bleus indique la quantité maximale d'information pouvant être extraire de notre jeu de données. Celle-ci correspond directement aux limites de représentation de notre conceptualisation, dont les valeurs ont été calculées en section 6.2.2. La ligne rouge continue représente la quantité d'information extraite par notre système. Les valeurs reportées sur les graphiques sont les F-scores calculés par la formule 6.6.

Le graphique 6.14a indique les valeurs des résultats obtenus pour la détection des cases, des bulles et des lignes de texte juste après la formulation des hypothèses de la première itération. L'apport du *système haut niveau* lors de la phase de validation de ces éléments est illustré par le graphique 6.14b. Le gain constaté est du même ordre de grandeur que celui de l'évaluation indépendante de chaque élément. Si la mesure de qualité de l'extraction des personnages reste relativement stable, le F-score des bulles et du texte est augmenté d'environ 3%.

Les graphiques 6.14c et 6.14d illustrent les résultats obtenus lors de la deuxième boucle du processus. Les régions d'intérêt des personnages sont estimées à partir des résultats obtenus lors de la première itération. Cette étape est logiquement celle qui souffre le plus de l'imperfection de l'extraction des éléments précédents. La propagation des erreurs successives, depuis les cases jusqu'à la détection de la position et de la direction de la flèche des bulles de dialogue, ramènent les résultats à un F-score de 11%. La baisse de 17% est significative mais compréhensible puisque



(a) Hypothèse (b) Validation et déduction

(c) Hypothèse (d) Validation et déduction

FIGURE 6.14 – Résultats de l'analyse continue des cases C, des bulles B, des lignes de texte T, des personnages P et de leur affiliation avec des bulles de dialogue BDP, contenus dans les planches de la base eBDtheque. La ligne en pointillés bleus représente l'objectif à atteindre, la ligne rouge représente les résultats. Les graphiques (a) et (b) illustrent les différentes étapes de la première itération, les graphiques (c) et (d) celles de la seconde.

les 28% obtenus précédemment étaient basés sur des informations fiables concernant les cases, les bulles et leurs flèches. La qualité de l'affiliation des personnages aux bulles de dialogue est ramenée en conséquence à un score de 18%.

6.5 Conclusion

Nous avons présenté dans ce chapitre un système complet pour l'analyse de planches de bandes dessinées. Il tire avantageusement partie de la conceptualisation proposée au chapitre 4 sur plusieurs points.

Il permet tout d'abord d'extraire, à partir des images, une partie des éléments visuellement complexes que sont les personnages. C'est une tâche particulièrement difficile car les personnages de bande dessinée peuvent prendre un nombre consi-

dérable de formes résultant en une grande hétérogénéité de leurs caractéristiques visuelles. Leur détection est effectuée sans apprentissage mais est guidée par la connaissance structurelle acquise progressivement sur le contenu de la planche. Il permet également d'augmenter légèrement les performances générales et très sensiblement la précision des résultats par son mécanisme de validation par contraintes.

Plusieurs pistes sont envisagées à court terme pour améliorer les performances de ce système. Nous avons vu que la propagation des imperfections dans l'extraction a un impact significatif sur la qualité des résultats produits en bout de chaîne. Afin de remédier à la baisse de rappel inhérente à la version actuelle, une possibilité serait, lors de l'étape de validation, de ne plus seulement rejeter les hypothèses mais, dans des cas de figure identifiés, de proposer un changement de classe. Une ligne de texte en contenant plusieurs autres pourrait par exemple être transformée en bulle.

Il est également possible d'ajouter des boucles d'interaction entre le *SHN* et le *SBN* pour compléter une information potentiellement manquante. Le cas d'une bulle validée par le modèle mais ne possédant pas de propriété la liant à une ligne de texte est un exemple parlant. De telles bulles pourraient être retournées au *système bas niveau* afin que celui-ci recherche d'éventuelles lignes de texte manquées.

Ces travaux ont été l'objet d'une publication soumise à *International Journal on Document Analysis and Recognition*, acceptée en novembre 2014.

Chapitre 7

Navigation et recherche d'informations

Sommaire

7.1 Introduction 125
7.2 Génération de la séquence de lecture 126
7.2.1 Problématique 126
7.2.2 Relations spatiales 126
7.2.3 Ordonnancement des cases 129
7.2.4 Intégration à l'ontologie 131
7.2.5 Évaluation 133
7.3 Contribution pour la recherche d'informations 135
7.3.1 Contexte 135
7.3.2 Navigation au sein d'un treillis 136
7.3.3 Optimisation du calcul des successeurs 140
7.3.4 Application et évaluation 142
7.4 Conclusion 146

7.1 Introduction

Le deuxième domaine d'intérêt identifié pour l'application de ces travaux de thèse concerne la recherche d'informations et la navigation à travers de larges bases de bandes dessinées annotées. La problématique est ici différente de celle déroulée dans le chapitre 6. Nous travaillions alors avec des données incertaines, dont nous tentions d'évaluer la valeur, et devant s'insérer dans une structure développée pour représenter les codes classiques de la bande dessinée. Nous nous plaçons à présent en aval de cette étape et partons du postulat que les informations présentes dans la base de connaissances ont été validées, corrigées, et qu'elles sont parfaitement fiables.

Ce dernier chapitre se décompose en deux sections bien distinctes. Nous présentons dans la première un mécanisme de génération de la séquence de lecture d'une planche. C'est une étape importante dans le processus d'optimisation de la lecture d'une bande dessinée numérique sur de petits écrans. La seconde section détaille des

travaux portant sur la navigation au sein d'un treillis de concepts, dans une perspective de recherche d'informations complexes au sein d'une base de bandes dessinées annotées.

7.2 Génération de la séquence de lecture

La première information que l'on peut chercher à faire émerger de cette structuration spatiale est l'ordre dans lequel doivent être lus les éléments d'une page. Nous nous intéresserons plus particulièrement aux cases, le principe pouvant être étendu d'une manière similaire aux bulles et aux lignes de texte.

7.2.1 Problématique

Comme nous l'avons vu précédemment, ces éléments sont organisés relativement les uns par rapport aux autres. L'ordre des cases est donc défini par rapport à la page, celui des bulles par rapport à la case les contenant et celui des lignes de texte par rapport à la bulle dans laquelle elles apparaissent.

Une case est traditionnellement de forme rectangulaire. Elle est séparée des autres cases par une bordure noire ou un espace intericonique de la couleur de fond de la page. Si l'auteur a recours à l'un ou l'autre selon la mise en page choisie, il aura à cœur, dans un souci de lisibilité de la planche, de rendre les différentes cases parfaitement identifiables. Un exemple mixte de séparation est présenté en figure 7.1.

Les cases ne sont pourtant pas toujours bien disjointes les unes des autres et alignées sur une grille de trois par trois. Il existe, notamment dans les mangas japonais mais aussi dans les comics et bandes dessinées européennes récentes, de nombreux exemples d'intersection spatiale entre plusieurs cases, se traduisant par un chevauchement partiel ou une inclusion complète d'une case dans une plus grande. La relation d'ordre entre deux cases n'est alors plus aussi nette que lorsque celles-ci sont franchement disjointes. Comme nous l'avons vu en section 3.2.1.2, l'auteur peut lui-même choisir d'intégrer ce genre d'ambiguïtés dans sa mise en page à des fins scénaristiques. Par exemple dans la figure 7.1, l'inclusion des cases inférieures dans la grande case permet d'insuffler à ces plans une impression de simultanéité temporelle des actions représentées. La plus grande case matérialisant une sorte d'arrière-plan à l'action, le lecteur est amené à promener son regard entre les petites cases et la grande pour apprécier la scène dans son ensemble.

7.2.2 Relations spatiales

Le concept *Panel* de l'ontologie $\mathcal{O}_{bd}$ étant subsumé par le concept *ROI* de $\mathcal{O}_{image}$, il est possible d'associer à chaque case ses coordonnées dans le repère de la page à travers l'attribut *hasWKT*. Nous allons alors formaliser les relations spatiales existantes entre les différentes instances de *Panel* partageant un lien avec une même instance de *Plate*. Nous considérons pour cette application que deux cases, C_i et

FIGURE 7.1 – Les deux cases en haut de la page sont séparées par une gouttière, ou espace intericonique, tandis que les deux cases tout en bas ne sont séparées de la case les englobant que par une bordure noire.

C_j entretiennent deux types de relations spatiales. Les premières sont des relations topologiques issues des huit relations du RCC-8 [Cohn 1997] (voir figure 7.2).

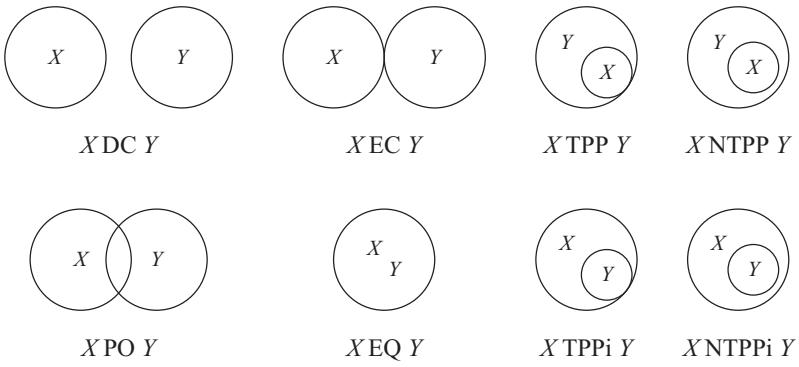


FIGURE 7.2 – Les huit relations du RCC-8.

Nous regroupons cet ensemble de relations selon trois situations observables sur une planche.

Soient A_i et A_j les intérieurs respectifs de C_i et C_j.

- $|A_i \cap A_j| = 0$, c'est-à-dire que C_i et C_j n'ont pas de surface commune.
  Cela correspond aux cas de disjonction (DC) et d'adjacence externe (EC). Nous considérons alors les cases comme étant séparables et nous notons cette relation SP.
- $0 < |A_i \cap A_j| < |A_i|, |A_j|$, c'est-à-dire que C_i et C_j partagent une surface de la page sans que l'une ne soit totalement incluse dans l'autre.
  Cela correspond au cas de chevauchement (PO).
- $A_i \cap A_j = A_i$ ou $A_i \cap A_j = A_j$, c'est-à-dire que C_i est incluse dans C_j ou inversement. Cela correspond aux cas d'inclusion tangentielle (TPP) et non tangentielle (NTPP) directes et inverses (TPPi et NTPPi). Nous considérons que l'une des cases contient ou est contenue par l'autre et nous notons cette relation CO.

Le deuxième type de relations spatiales entretenues entre deux cases est celui de l'orientation cardinale. Nous considérons qu'une case C_i peut être au-dessus (Nord), à droite (Est), au-dessous (Sud) ou à gauche (Ouest) d'une case C_j. Le cerveau humain ordonnant naturellement les cases horizontalement puis verticalement, l'introduction des orientations intermédiaires de type Nord-Est ajouterait plus de confusion que de précision. La position relative de C_i par rapport à C_j est déterminée d'une manière différente selon la relation topologique qu'elles entretiennent. Disons que l'on cherche à déterminer la véracité de la relation $Sud(C_i, C_j)$, c'est-à-dire le fait que la case C_i soit sous la case C_j. Si les deux cases sont séparables, tous les points de C_j devront avoir une ordonnée supérieure à tous les points de C_i pour que la relation soit validée (voir figure 7.3a et équation 7.1). De même, pour valider la relation $Est(C_i, C_j)$, les abscisses de C_j devront être supérieures à celles de C_i (voir figure 7.3b).

$$Sud(C_i, C_j) \Leftarrow SP(C_i, C_j),\ (\forall(y_{i_i}, y_{j_i}),\ y_{j_i} > y_{i_i}) \tag{7.1}$$

Dans le cas où deux cases présentent une intersection, nous les réduisons à leur centre de gravité et nous comparons leur position selon un modèle en cône à quatre divisions. La relation $Sud(C_i, C_j)$ est vérifiée si le centre de gravité de C_j se trouve dans le cône compris entre $-\pi/4$ et $-3\pi/4$ sur la représentation standard du cercle trigonométrique et ayant pour origine le centre de C_i (voir figure 7.3c et équation 7.2). La relation Est est calculée de manière similaire pour un angle compris entre $-\pi/4$ et $\pi/4$. Des équations 7.2 et 7.1 découle que $Sud(C_i, C_j)$ et $Est(C_i, C_j)$ sont inversibles et que $Nord(C_i, C_j) \Leftrightarrow Sud(C_j, C_i)$ et $Ouest(C_i, C_j) \Leftrightarrow Est(C_j, C_i)$. Enfin, dans le cas où l'on a $CO(C_i, C_j)$, c'est-à-dire lorsque la case C_i est contenue dans la case C_j ou inversement, nous considérons que la relation d'ordre est indécidable. Cela dans le but d'intégrer la part d'interprétation du lecteur souhaitée par l'auteur à l'ordre de lecture généré.

$$Sud(C_i, C_j) \Leftarrow PO(C_i, C_j),\ (y_{j_c} > y_{i_c}),\ (y_{j_c} - y_{i_c} > |x_{j_c} - x_{i_c}|) \tag{7.2}$$

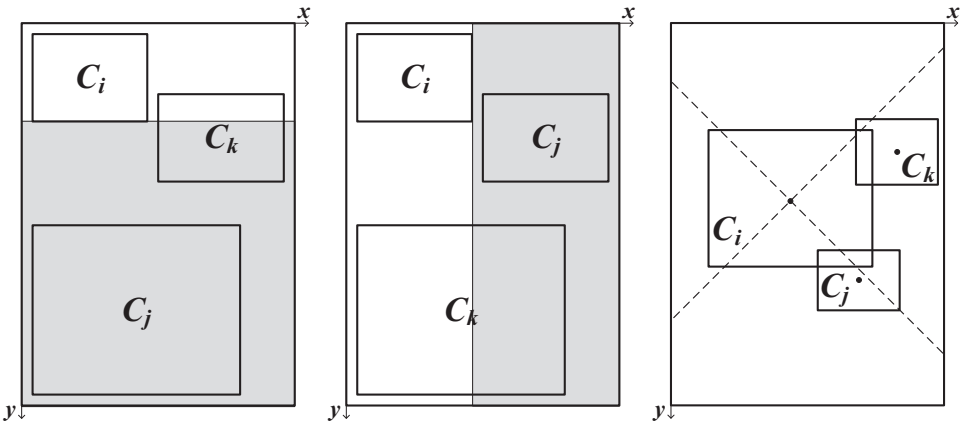


FIGURE 7.3 – (a) C_i, C_j et C_k sont séparables, C_j est au sud de C_i mais pas C_k. (b) De la même manière, C_j est à l'est de C_i mais pas C_k. (c) C_i, possède une intersection avec C_j et C_k, C_j est au sud de C_i mais C_k est à l'est

7.2.3 Ordonnancement des cases

Grâce à la définition des prédicats des relations spatiales topologiques et cardinales considérées entre deux cases, nous pouvons maintenant poser la formule qui va nous permettre de déterminer l'ordre dans lequel peut être lu un ensemble de cases sur une page donnée. Soit la relation $Suivant(C_i, C_j)$ exprimant le fait que la case C_j doit se lire plus tard que la case C_i. Pas nécessairement *juste après* C_i mais plus tard. Cette relation est valide dans les cas où C_j est au sud ou bien au même niveau et à l'est de C_i, tel que le formalisent les équations 7.3 et 7.4.

$$Suivant(C_i, C_j) \Leftarrow Sud(C_i, C_j) \tag{7.3}$$

$$Suivant(C_i, C_j) \Leftarrow \neg Sud(C_i, C_j),\ \neg Sud(C_j, C_i),\ Est(C_i, C_j) \tag{7.4}$$

A noter que l'équation 7.4 est valide pour le paradigme de lecture occidental allant de la gauche vers la droite. Si l'on considère le cas des mangas dans lesquels les planches se lisent traditionnellement de droite à gauche, les éléments de la relation *Est* sont inversés et l'équation devient :

$$Suivant(C_i, C_j) \Leftarrow \neg Sud(C_i, C_j),\ \neg Sud(C_j, C_i),\ Est(C_j, C_i) \tag{7.5}$$

Une case C_i est alors en relation avec un ensemble S_i de cases, dont la cardinalité est comprise entre 0 et n, via la relation *Suivant*. L'ordre de lecture est généré grâce à un algorithme de tri fusion appliqué sur l'ensemble des cases d'une planche. La fonction de comparaison de deux éléments de l'ensemble, déterminant si l'un doit être classé avant ou après l'autre, est décrite par l'algorithme 1. Elle peut retourner trois valeurs pour un couple de cases C_i et C_j, en fonction de S_i et S_j. Si $C_j \in S_i$

Nom : `compare`
Entrées : Deux cases C_i et C_j et leurs ensemble de successeurs respectifs S_i et S_j.
Sorties : Un entier compris entre -1 et 1 indiquant la position de C_i par rapport à C_j.

```
début
    si CO(C_i, C_j) alors
        retourner 0
    fin
    sinon si C_j ∈ S_i alors
        retourner -1
    fin
    sinon
        retourner 1
    fin
fin
```

Algorithme 1 : Fonction de comparaison de deux cases pour leur ordonnancement séquentiel

alors C_i est classée avant C_j, sinon si $C_i \in S_j$ alors C_i est classée après C_j, sinon C_i et C_j sont classées au même rang.

Il faut noter que, contrairement à la notion d'ordre de lecture, la fonction `compare` n'est pas transitive. C'est-à-dire que si l'on a `compare`$(C_i, C_j) = -1$ et `compare`$(C_j, C_k) = -1$, nous n'avons pas pour autant la certitude que l'égalité `compare`$(C_i, C_k) = -1$ soit vérifiée.

En effet, si l'on s'intéresse au cas d'exemple présenté en figure 7.4, selon l'équation 7.4, les cases hachurées horizontalement doivent être lues avant la case quadrillée. Toujours selon 7.4, cette dernière doit elle-même être lue avant les cases hachurées verticalement. Cependant, selon la relation 7.3, les cases hachurées horizontalement doivent être lues après les cases hachurées verticalement. Ces dernières doivent donc, paradoxalement, être lues à la fois avant et après la case quadrillée.

L'algorithme de tri par fusion étant stable, l'ordre final des cases dépendra, pour les cas ambigus, de l'ordre dans lequel l'ensemble de départ a été fourni à l'algorithme.

Ces cas de figure particuliers sont néanmoins aisément détectables en vérifiant que, pour toute case C_i d'indice i et pour toute case C_l d'indice l avec $l > i$, on a $C_l \in S_i$. Cette mesure englobe également les cas où deux cases peuvent être lues dans n'importe quel ordre, mais ceux-ci peuvent être filtrés car leur comparaison selon l'algorithme 1 retourne 0.

Une fois l'incohérence détectée, une intervention humaine peut être demandée afin de préciser l'ordre de lecture correct. Comme le soulignent les expérimentations détaillées en section 7.2.5, ces cas de figures sont marginaux et n'apparaissent que sur un nombre extrêmement limité de pages. Nous pouvons toutefois noter qu'un comportement similaire peut être observé dans le cas où la planche a été positionnée

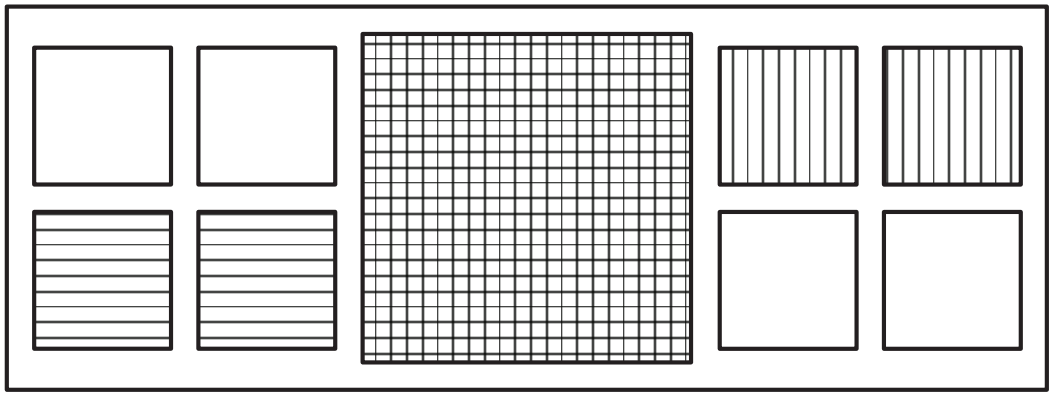


FIGURE 7.4 – Cas particulier produisant un ordre de lecture incohérent.

légèrement en biais lors de sa numérisation. Le repère orthonormé de l'image correspondant alors à celui de la planche modulo un léger angle de rotation. Deux cases, distantes et disjointes en ordonnées dans le repère de la page, pourront alors avoir une intersection en ordonnées dans le repère de l'image. C'est pourquoi en pratique nous introduisons un seuil de tolérance proportionnel à un angle de rotation de un degré.

7.2.4 Intégration à l'ontologie

Cette méthode de tri fournit un ordre de lecture partiel (voir figure 7.6b) que l'on peut soit dériver en ordre total via un tri topologique, soit intégrer tel quel à la base de connaissances. La place établie i pour chaque case C dans l'ordre choisi est formalisée à travers l'assertion d'un axiome *hasRank(C, i)* dans la A-Box de l'ontologie. Dans le cas où l'incertitude de la relation d'ordre entre plusieurs cases a été conservée, celles-ci se voient attribuer la même valeur de rang. Une case contenant plusieurs cases plus petites aura alors autant d'instances d'attribut *hasRank* que de cases contenues (voir figure 7.5).

Une nouvelle règle 7.6 est ajoutée à l'ontologie $\mathcal{O}_{bd}$ pour formaliser les relations de successions entre ces cases.

$$\begin{aligned} &\text{Plate}(\,?\text{p}),\ \text{Extractor}(\,?\text{e}),\\ &\text{Panel}(\,?\text{p1}),\ \text{Panel}(\,?\text{p2}),\\ &\text{hasPanel}(\,?\text{p},?\text{p1}),\ \text{hasPanel}(\,?\text{p},?\text{p2}),\\ &\text{hasExtractor}(\,?\text{p1},?\text{e}),\ \text{hasExtractor}(\,?\text{p2},?\text{e}),\\ &\text{hasRank}(\,?\text{p1},?\text{r1}),\ \text{hasRank}(\,?\text{p2},?\text{r2}),\\ &\qquad\qquad \text{equal}(\,?\text{r1},?\text{r2}) \rightarrow \text{hasNextPanel}(\,?\text{p1},?\text{p2}) \end{aligned} \tag{7.6}$$

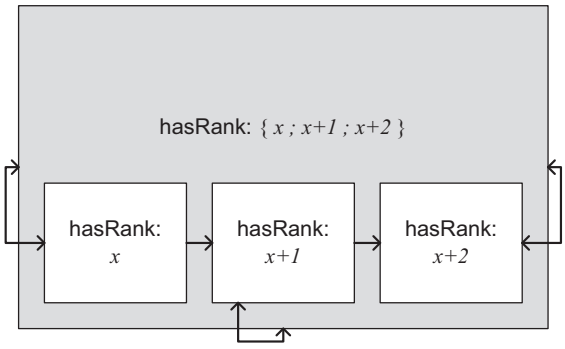


FIGURE 7.5 – Relations dans le cas d'une inclusion multiple entre une case contenante (en gris) et un ensemble de cases contenues (en blanc). Les flèches représentent les relations *hasNextPanel* établies entre les cases.

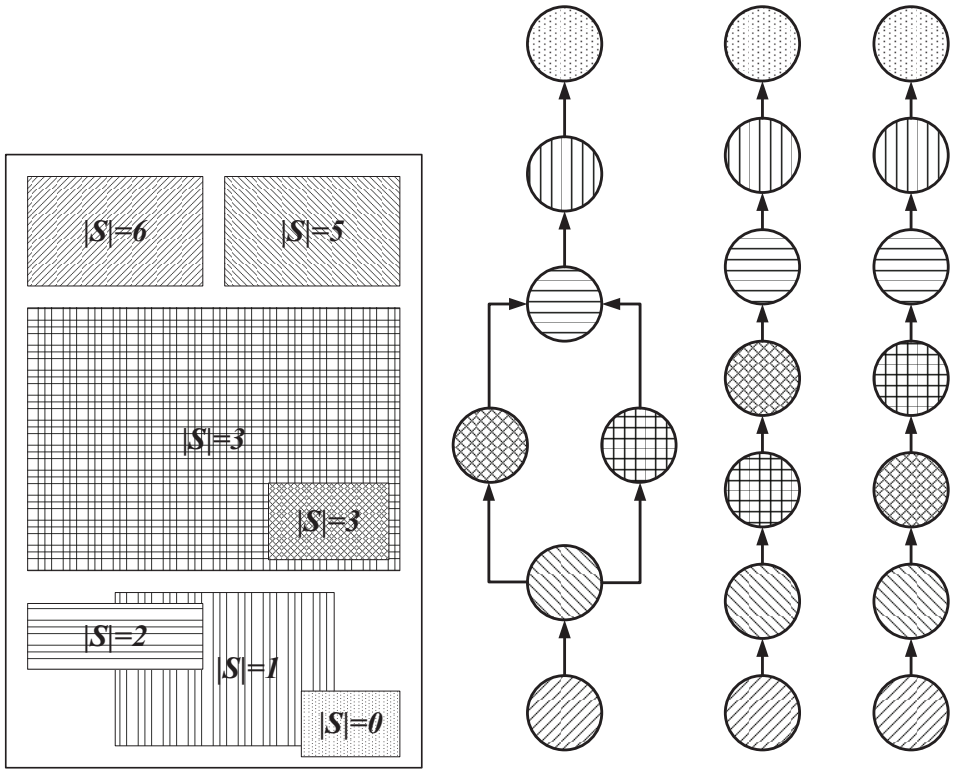


FIGURE 7.6 – (a) Un exemple de contenu d'une page. S représente l'ensemble des cases devant être lues après une case donnée selon l'équation 7.3. (b) L'ordre partiel qui découle du tri de l'ensemble des cases selon la fonction de comparaison 1 . (c) Les deux ordres totaux pouvant être générés par un tri topologique de l'ordre partiel.

7.2.5 Évaluation

Nous décrivons ici l'expérimentation, et son évaluation, de l'algorithme de génération de l'ordre de lecture des cases au sein d'une planche de bande dessinée.

Nous avons pu évaluer la génération de l'ordre de lecture des cases d'une page grâce à la vérité terrain présentée dans le chapitre 5.

Soient $r_i = r(C_i)$ la fonction associant un rang r_i à une case C_i dans la vérité terrain et $r'_i = r'(C_i)$ la fonction associant un rang r'_i à une case C_i via notre système de tri. Pour chaque couple de cases $\{C_i, C_j\}$ tel que $r'_i = r'_j + 1$, nous évaluons la véracité de l'égalité formulée par l'équation 7.7.

$$r(C_i) - r(C_j) = r'(C_i) - r'(C_j) = 1 \tag{7.7}$$

Si deux cases C_i et C_j adjacentes dans l'ordre de lecture calculé ne le sont pas dans la vérité terrain, l'égalité n'est pas vérifiée et la relation calculée est considérée comme erronée.

Cette mesure est moins punitive mais plus juste que la comparaison directe de r_i et r'_i pour chaque C_i car elle permet d'évaluer la position relative des cases les unes par rapport aux autres dans l'ordre de lecture, plutôt que leur position absolue. Ainsi, une erreur en début de séquence ne sera sanctionnée qu'une seule fois, au lieu d'être reportée en cascade sur l'ensemble des cases de la page.

Sur les 849 cases réparties sur les 100 pages que compte la base eBDtheque, 749 relations de succession peuvent être évaluées. Les résultats de ces évaluations sont présentés dans le tableau 7.1.

Base de test	Nombre de relations	Résultats
Base entière (BE)	749	94,65%
BE - Inclusions (I)	746	95,03%
BE - I - Doubles pages (DP)	703	97,58%
BE - I - DP - Page extrême (PE)	581	99%

TABLEAU 7.1 – Résultats de notre méthode d'ordonnancement des cases

Les résultats sur la base entière sont bons, avec un taux de validité supérieur à 94%, qui peut être mis en perspective avec la source des erreurs constatées.

Premièrement, la base ne contient que des ordres de lecture totaux, interprétation suggestive des annotateurs, et ne gère donc pas la simultanéité de lecture des cases incrustées. Notre algorithme attribuant le même rang aux cases incluses les unes dans les autres, l'égalité de l'équation 7.7 n'est pas vérifiée dans ces cas là. Pour autant, l'ordre de lecture établi dans la vérité terrain correspond bien à l'un des ordres de lecture générés par un tri topologique de l'ordre partiel. La prise en compte de ce paramètre fait monter les résultats de 0.5%.

Deuxièmement, la base eBDtheque contient cinq images représentant des doubles pages contenant deux planches individuelles. C'est-à-dire que les cases présentes sur

ces images se sont pas sensées se parcourir sur une double page mais doivent se lire classiquement, d'abord sur la page de gauche, ensuite sur la page de droite. Notre algorithme a été conçu pour fonctionner planche par planche, il est par conséquent peu pertinent d'évaluer ses performances sur de tels exemples. Les retirer du jeu de données porte le taux de bonnes affectations à plus de 97%.

Enfin, la troisième source d'erreurs provient de la planche unique, à la mise en page extrême et au nombre de cases hors du commun, présentée en figure 7.7. Celle-ci met en exergue la faiblesse identifiée précédemment, et illustrée par la figure 7.4, dont on retrouve le schéma à plusieurs endroits sur la planche.

Notre méthode se positionne honorablement en termes de performances et peut être comparée à celle détaillée dans [Ponsard 2012]. L'approche développée dans ces travaux présente des points communs avec la nôtre, notamment sur la manière de gérer la disjonction et le chevauchement des cases. Elle diffère cependant sur le comportement adopté en cas d'inclusion, la grande case étant ordonnée avant la case incluse. Notre approche autorise dans ce cas de figure le lecteur à naviguer entre une case et son contenant autant de fois qu'il le désire à travers la propriété *hasNextPanel*. Enfin, la méthode [Ponsard 2012] ne permet pas non plus de gérer les exceptions du type de celle illustrée en figure 7.4. Les expérimentations ayant été menées sur des jeux de données différents, il est malheureusement impossible d'effectuer une comparaison plus formelle entre nos deux méthodes.

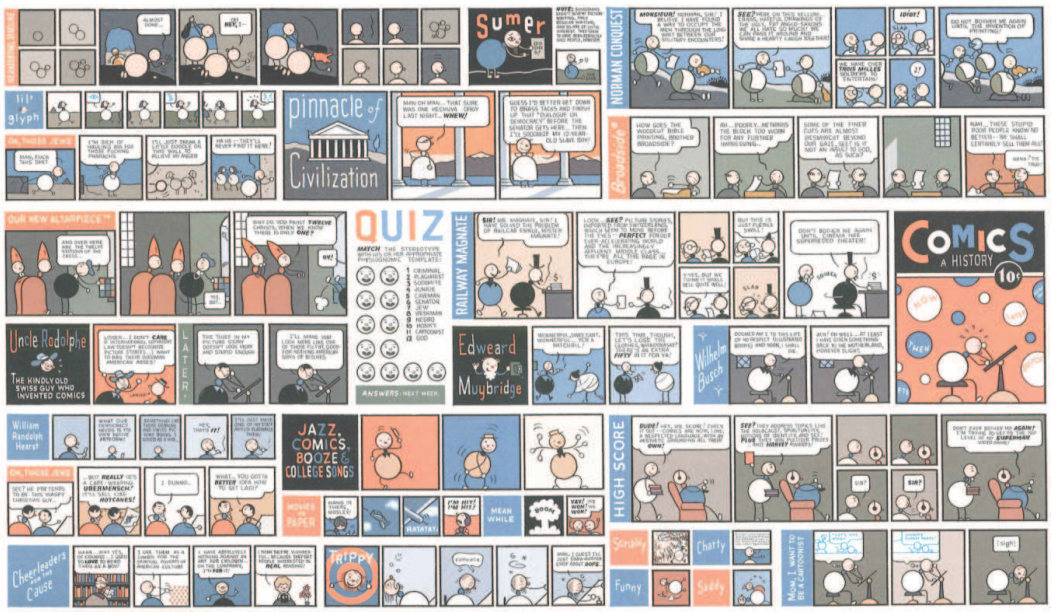


FIGURE 7.7 – Exemple de mise en page extrême. Crédits : Chris Ware - Acme Novelty Library - New York Pantheon

Une partie de ces travaux a été présentée à l'occasion des 23ème journées francophones d'ingénierie des connaissances.

7.3 Contribution pour la recherche d'informations

7.3.1 Contexte

Il existe de nombreuses manières de rechercher de l'information dans un ensemble de données. Le choix d'une méthode plutôt qu'une autre dépend notamment de la structure dans laquelle ces données s'intègrent, cette structure dépendant elle-même du type des données stockées.

Quoi qu'il en soit, la recherche d'information a toujours eu à faire face à des compromis. La recherche par mots clés doit par exemple faire face au problème de l'hétérogénéité sémantique exposé en section 3.5.1. Il n'y a non seulement aucune garantie que les résultats retournés correspondent tous à ce que recherche l'utilisateur, mais leur complétude n'est pas non plus assurée. Tous les éléments correspondant à la requête d'un utilisateur ne font en effet pas nécessairement partie de l'ensemble retourné par le système.

Les langages de requête tels que SQL ou SPARQL apportent une solution à ce problème. A chaque requête correspond un ensemble réponse comprenant uniquement les éléments, et tous les éléments, conformes aux caractéristiques demandées, pour peu que la requête ait été bien formulée. C'est bien d'ailleurs l'inconvénient de ces langages. Ils ne sont pas intuitifs pour un utilisateur lambda et le vocabulaire formel décrivant les individus et leurs relations dans la base de données ou dans la base de connaissances doit être connu en amont. Une petite erreur dans la requête peut mener à une absence de résultats, ou pire, à l'affichage de résultats ne correspondant pas au souhait de l'utilisateur.

Les systèmes de recherche par navigation, tels que les arborescences de dossiers et de fichiers, ne partagent pas ce problème. Leur utilisation est tout à fait intuitive et ne nécessite pas de connaissances a priori sur le vocabulaire employé pour structurer les données. Cette structure a en revanche l'inconvénient d'être fixe, ce qui implique des restrictions dans les approches pouvant être employées pour trouver des éléments. Si une collection de photos est classée dans une arborescence de dossiers basée sur les dates, il est impossible de rechercher, à travers cette arborescence, toutes les photos prises au même endroit ou celles figurant les mêmes personnes.

Plus récemment sont apparus les systèmes de recherche d'informations par facettes [Allen 1996]. Ces systèmes permettent d'élaguer l'espace de recherche en apposant des filtres successifs sur un ensemble de résultats. On retrouve par exemple un tel système dans Google Images qui propose de filtrer les résultats par date, couleur, type de contenu, etc. Le problème des systèmes à facettes est qu'ils ne sont pas des systèmes de recherche d'informations à part entière. Ils sont plus généralement utilisés en complément d'un système de recherche par mots clés, permettant d'affiner les requêtes SQL renvoyant les résultats souhaités.

Parallèlement, les structures de données basées sur les treillis [Birkhoff 1967] ont commencé à être un objet d'attention pour la recherche d'informations [Godin 1993]. Des problèmes de performances en termes de temps de calcul ont cependant freiné

leur développement jusqu'au début des années 2000 et aux travaux de [Godin 1989, Carpineto 1995, Priss 2000]. Récemment, des systèmes d'information logiques (LIS) mêlant les capacités d'organisation des données des treillis, le pouvoir de représentation des connaissances apporté par les logiques de description et les mécanismes de recherche par facettes ont émergé [Ferré 2002].

De telles approches peuvent être mises à profit pour la recherche d'informations de manière complète (certitude d'obtenir tous les résultats correspondant à une requête) et sûre (impossibilité de formuler une requête ne renvoyant aucun résultat) [Ferré 2011]. Elle ont également été utilisées pour l'annotation manuelle d'images, et notamment d'images de bandes dessinées [Hermann 2012], à travers l'application Sewelis (voir figure 7.8)

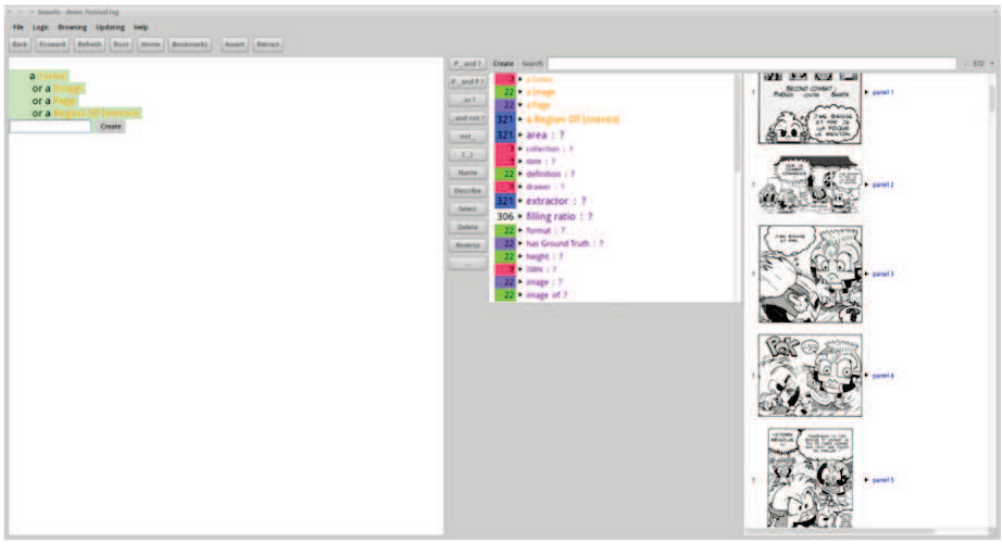


FIGURE 7.8 – Interface de requêtage par navigation de Sewelis

Les problèmes computationels liés au parcours de treillis, s'ils ont tendance à s'amenuiser avec le temps, restent toutefois l'un des points épineux plaçant ces systèmes en retrait lorsqu'ils sont comparés, par exemple, aux bases de données relationnelles ou aux entrepôts de triplets. Dès lors, la question de l'optimisation de la navigation au sein d'un treillis se pose.

7.3.2 Navigation au sein d'un treillis

7.3.2.1 Définition et génération d'un treillis de concepts

Un treillis de concepts se défini de manière formelle à partir d'une table binaire, aussi appelé contexte formel, $(O, I, (\alpha, \beta))$ où O représente l'ensemble des individus (aussi appelés observations ou objets), et I l'ensemble de leurs caractéristiques (ou attributs). $\alpha(A)$ est l'ensemble d'attributs partagés par tous les membres d'un

sous-ensemble A d'objets. Réciproquement, $\beta(B)$ est l'ensemble des observations partageant le même sous-ensemble de caractéristiques B. Chaque nœud d'un treillis de concepts est un concept (A, B), c'est-à-dire la correspondance maximale entre les objets et les attributs vérifiant $\alpha(A) = B$ et $\beta(B) = A$. On appelle A l'extension du concept et B l'intension.

Deux concepts formels (A_1, B_1) and (A_2, B_2) sont reliés dans le treillis lorsqu'ils vérifient la propriété de spécialisation-généralisation suivante :

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow \left\| \begin{array}{l} A_1 \subseteq A_2 \\ \text{(équivalent à } B_1 \supseteq B_2) \end{array} \right.$$

La structure formée par l'ensemble des concepts formels vérifiant la relation d'ordre $\leq$ est appelée treillis des concepts, ou treillis de Galois. La relation de couverture $\prec$ forme le diagramme de Hasse du treillis. Le concept absurde ($\perp$) regroupe l'intégralité des attributs associés aux individus les partageant, généralement aucun. Le concept général ($\top$) contient à l'inverse l'ensemble des observations et les attributs qu'ils partagent. Chaque élément du treillis peut être vu comme une généralisation ou une spécialisation d'un concept, la treillis entier étant alors assimilable à une hiérarchie de concepts [Wille 1982]. Cette structure hiérarchique semble particulièrement adaptée pour la recherche d'informations, les concepts la composant étant à la fois sémantiquement et spatialement proches.

La lecture du livre de référence écrit par Ganter et Wille [Ganter 1999] est conseillée pour une compréhension complète des treillis de concepts.

De nombreux algorithmes ont été proposés dans la littérature pour la génération des treillis [Norris 1978, Ganter 1984, Bordat 1986, Nourine 1999]. Tous les algorithmes de construction proposés sont d'une complexité polynomiale en fonction du nombre de concepts (au mieux quadratique pour [Nourine 1999]). Les dimensions du treillis, c'est-à-dire le nombre de concepts qui le composent, déterminent directement la complexité de sa génération. Le nombre de concepts étant exactement compris entre $2^{|O+I|}$ dans le pire des cas, c'est-à-dire quand la table binaire est une matrice diagonale de zéros, et entre $|O + I|$ dans le meilleur des cas (diagonale de uns). Une étude comparative de ces différents algorithmes a été menée par [Kuznetsov 2002], concluant sur des adéquations différentes entre certains algorithmes et les caractéristiques des données à traiter, plutôt que sur une supériorité nette d'une méthode par rapport aux autres. En effet, bien que ces algorithmes produisent tous en sortie le même treillis, les approches employées sont différentes. L'algorithme de référence *Next Closure* [Ganter 1984] détermine l'ordre lectique des concepts, ceux-ci devant ensuite être ordonnés par $\leq$ où $\prec$ pour former le treillis. L'algorithme de Bordat [Bordat 1986] a été le premier algorithme à générer directement le diagramme de Hasse du treillis. Les successeurs immédiats de chaque concepts sont calculés, depuis le concept absurde jusqu'au concept général. La possibilité de calcul des successeurs immédiats d'un concept est très intéressante pour une navigation localisée à l'intérieur de la structure du treillis. Elle ne suppose pas la génération de l'ensemble

des concepts, ce qui représente un gain, en termes de temps de calcul et d'espace, pouvant être considérable.

7.3.2.2 Calcul des successeurs immédiats d'un concept

L'algorithme de Bordat, redécouvert de manière indépendante par [Lindig 2002] est dérivé d'un corollaire du théorème de Bordat [Bordat 1986] posant qu'un concept (A', B') est un successeur immédiat d'un concept (A, B) si et seulement si A' est un sous-ensemble inclus-maximal dans un système $\mathcal{F}_A$ défini sur l'ensemble O des observations par la formule 7.8.

$$\mathcal{F}_A = \{\beta(x) \cap A \ : \ x \in I \backslash B\} \tag{7.8}$$

Exemple : Soit $A = \{a, b, c, d, e, f\}$ un ensemble et $A_1 = \{a, c, d, e, f\}$, $A_2 = \{c, b\}$ et $A_3 = \{a, d, e\}$ des sous-ensembles de A. A_1 et A_2 sont des sous-ensembles inclus-maximaux de A car leurs éléments sont inclus dans A et qu'ils ne sont eux-mêmes pas inclus dans d'autres sous-ensembles de A. A_3 en revanche n'est pas un sous-ensemble inclus-maximal car il est inclus dans A_1.

L'algorithme génère tout d'abord le système d'ensembles $\mathcal{F}_A$ en un temps linéaire. Les sous-ensembles inclus-maximaux sont ensuite calculés en $\mathcal{O}(|O|^3)$ via l'utilisation, par exemple, d'un graphe d'inclusion. On peut noter que le problème du calcul des sous-ensemble inclus-minimaux est connu pour pouvoir être résolu en $\mathcal{O}(|O|^{2,5})$ via l'emploi de structures de données plus sophistiquées [Fischer 1971, Munro 1971].

Il est possible de travailler avec un treillis réduit à son seul ensemble d'attributs (ou d'observations). Il a en effet été établi que tout treillis de concepts est isomorphe à un treillis des fermés sur l'ensemble I des attributs (respectivement O). Chaque concept du treillis étant alors seulement composé d'attributs, sans information supplémentaire sur les objets correspondants (ou inversement), ce qui peut, selon les cas d'utilisation, donner lieu à un gain de performances important en termes de stockage et de temps de calcul. Le treillis des fermés est constitué de toutes les fermetures calculées par l'opérateur $\varphi = \alpha \circ \beta$.

Exemple : Soit les concepts $(\{1\}; \{a, c, f\})$, $(\{2\}; \{b, c, d, f\})$ et $(\{3\}; \{b, e, f\})$, le fermé généré à partir de l'attribut c sera égal à :

$$\begin{aligned} \varphi(c) &= \alpha \circ \beta(\{c\}) \\ \varphi(c) &= \alpha(\{1, 2\}) \\ \varphi(c) &= (c, f) \end{aligned}$$

Un panorama dressé par [Caspard 2003] est conseillé au lecteur désirant plus de détails sur les treillis des fermés.

Une extension de l'algorithme de Bordat permet de générer le treillis des fermés. Chaque successeur immédiat B' d'un fermé B est obtenu par $B' = \alpha(A')$ avec

$A' \in \mathcal{F}_A$. Or $A' = \beta(x) \cap A = \beta(x) \cap \beta(B) = \beta(x + B)$, donc $B' = \alpha(A') = \alpha(\beta(x+B)) = \varphi(x+B)$. Les successeurs immédiats d'un fermé B sont donc les sous-ensembles inclus-minimaux, dans un système $\mathcal{F}_B$ défini sur l'ensemble des attributs I par la formule 7.9.

$$\mathcal{F}_B = \{\varphi(x + B) \ : \ x \in I \backslash B\} \tag{7.9}$$

L'algorithme 2 déroule le protocole de calcul des successeurs immédiats d'un treillis des fermés.

Nom : `Successeurs_Immediats`
Entrées : Un contexte $K = (O, I, (\alpha, \beta))$; Un fermé B du treillis de K
Sorties : Les successeurs immédiats de B dans le treillis des fermés
début
  Initialisation du système d'ensembles $\mathcal{F}_B$ par $\emptyset$;
  pour chaque $x \in I \backslash B$ **faire**
    Ajouter $\varphi(x + B)$ à $\mathcal{F}_B$
  fin
  $Succ$=sous-ensembles inclus-minimaux de $\mathcal{F}_B$;
  return $Succ$
fin

Algorithme 2 : Algorithme de calcul des successeurs immédiats

7.3.2.3 Limited Object Access

L'algorithme Limited Object Access (LOA), présenté par Demko et Bertet en 2011 [Demko 2011] est une autre extension des successeurs immédiats de Bordat. L'idée apportée au sein de cet algorithme est inspirée des bases de données relationnelles et consiste en un comptage des objets. Au lieu de considérer l'extension d'un sous-ensemble d'attributs comme un sous-ensemble d'observations, le calcul des inclus-maximaux de $\mathcal{F}_A$ est effectué en ne tenant compte que de la cardinalité des extensions. Pour chaque attribut $x \in I \backslash B$ et chaque successeur potentiel identifié $X \subseteq I \backslash B$, quatre cas de figure sont relevés, dépendants de la relation d'inclusion existant entre $\varphi(B + X)$ et $\varphi(B + x)$.

1. Fusionner x avec X lorsque $\varphi(B + x) = \varphi(B + X)$
2. Supprimer x de la liste des successeurs potentiels de B lorsque $\varphi(B + x) \supset \varphi(B + X)$
3. Supprimer X de la liste des successeurs potentiels de B lorsque $\varphi(B + x) \subset \varphi(B + X)$
4. Insérer x dans la liste des successeurs potentiels de B lorsque x n'est ni éliminé, ni fusionné avec X

Le test d'inclusion entre les ensembles $\varphi(B + X)$ et $\varphi(B + x)$ peut être effectué très rapidement grâce à une fonction de comptage c, dérivée de la proposition de [Demko 2011].

Proposition 1 ([Demko 2011]) :

$$\varphi(B+X) \subseteq \varphi(B+x) \Longleftrightarrow c(B+X+x) = c(B+X) \tag{7.10}$$

La fonction de comptage c associe à tout ensemble d'attributs X la cardinalité de son extension $\beta(X)$.

$$c(X) = |\beta(X)| \tag{7.11}$$

$$\varphi(B) = B + \{x \in I \backslash B \ : \ c(B) = c(B+x)\} \tag{7.12}$$

Cette fonction de comptage renvoie à la notion de support introduite dans la recherche de règles d'association, particulièrement utilisé par l'algorithme Titanic [Stumme 2002].

L'algorithme 3 propose une retranscription complète de Limited Object Access.

7.3.3 Optimisation du calcul des successeurs

Les performances de l'algorithme LOA sont directement liées à l'efficacité de la fonction de comptage des éléments. Les temps de calcul mis en avant au sein de l'article de présentation de l'algorithme tirent partie des capacités d'indexation et de comptage du moteur SQL. Nous proposons ici sa généralisation à des langages de programmation indépendants de toute base de données.

En observant l'exécution d'une implémentation naïve de l'algorithme, nous avons en effet constaté que 86% du temps de calcul était alloué au calcul de l'extension des différents ensembles d'attributs candidats $\beta(X)$. Or, l'extension d'un ensemble d'attributs est égal à l'intersection des extensions de ses éléments individuels.

$$\beta(X) = \bigcap_{i=1}^{|X|} \beta(x_i) : x \in X$$

Dans le processus de génération des extensions, la tâche la plus chronophage est justement l'opération d'intersection entre deux ensembles. Elle représente en effet 87% du calcul des extensions soit près des trois quarts du temps d'exécution total. Si l'opération en elle-même ne prend qu'environ cinquante microsecondes sur une machine modeste, la facteur multiplicatif des occurrences d'appel grandit rapidement en fonction de la taille du contexte, résultant en un temps d'exécution pouvant dépasser plusieurs minutes. Une attention toute particulière doit donc être portée à l'optimisation du calcul des intersections, à la fois en termes de nombre d'appels et en temps de traitement.

Le nombre d'appels de la fonction de comptage peut premièrement se limiter à trois à l'intérieur de chaque boucle « *pour chaque* ». En effet, les valeurs de $c(B+x)$, $c(B+X)$ and $c(B+X+x)$ sont stables tout au long de la deuxième boucle, elles peuvent donc se calculer une fois pour toutes en amont de l'exécution de cette dernière. Dans le cas contraire, le nombre d'appels aurait pu osciller entre quatre et huit (deux pour chaque condition « *si* » et deux supplémentaires pour l'une d'entre elles).

Name : `Successeurs_Immediats_LOA`
Entrées : Un contexte $K = (O; I, (\alpha, \beta))$; Un fermé B du treillis des fermés $(\mathbb{C}_I, \subseteq)$ of K
Sorties : Les successeurs immédiats de B dans le treillis
début
  Initialisation du système d'ensembles $Succ_B$ avec $\emptyset$;
  pour chaque $x \in I \setminus B$ **faire**
    add = true;
    pour chaque $X \in Succ_B$ **faire**
      \\ **Cas 1 : Fusionner x et X** en un successeur potentiel unique
      si $c(B + x) = c(B + X)$ **alors**
        si $c(B + X + x) = c(B + x)$ **alors**
          remplacer X par $X + x$ dans $Succ_B$;
          add=false ; break;
        fin
      fin
      \\ **Cas 2 : Supprimer x** de la liste des successeurs potentiels
      si $c(B + x) < c(B + X)$ **alors**
        si $c(B + X + x) = c(B + x)$ **alors**
          add=false ; break;
        fin
      fin
      \\ **Cas 3 : Supprimer X** de la liste des successeurs potentiels
      si $c(B + x) > c(B + X)$ **alors**
        si $c(B + X + x) = c(B + X)$ **alors**
          supprimer X de $Succ_B$
        fin
      fin
    fin
    \\ **Cas 4 : Insérer x** dans la liste des successeurs potentiels ;
    si *add* **alors**
      ajouter $\{x\}$ à $Succ_B$
    fin
  fin
  retourner $Succ_B$;
fin

Algorithme 3 : Algorithme de calcul des successeurs immédiats par la méthode Limited Object Access.

Les structures de données fournies par les langages de programmation, tels que les *HashSet* ou les *TreeSet* en Java, sont particulièrement adaptés au stockage et à l'organisation des objets, de leurs attributs et de leurs intensions et extensions respectives. Ils semblent en revanche trop lourds pour la représentation d'un simple

contexte sous la forme d'une table binaire. En réalité, les objets n'ont pas nécessairement besoin d'être manipulés directement pour calculer l'extension, et a fortiori la cardinalité de l'extension, d'un ensemble d'attributs.

Si l'on suppose que l'ensemble O des observations est trié dans un ordre strict et immuable, chacune d'entres elles peut être représentée et manipulée par son indice au sein de cet ordre. L'extension d'un attribut devient alors un mot binaire dont la longueur est égale à la cardinalité de l'ensemble des observations, $|O|$. Naturellement, les éléments de ce mot peuvent prendre la valeur 1 ou 0, indiquant respectivement la présence ou l'absence, au sein de l'extension de l'attribut, de l'observation d'indice correspondant. Il existe de nombreux moyens de manipuler des mots binaires, Java propose par exemple sa propre classe *BitSet* et les méthodes de manipulation associées.

L'extension (respectivement l'intension) de chaque attribut (resp. observation) du contexte est calculée et stockée sous la forme d'un mot binaire, une fois pour toutes, au début de l'exécution du programme. A partir de là, l'extension d'un ensemble d'attributs peut être calculée via l'application d'un *ET logique* sur l'ensemble d'extensions correspondant (voir l'exemple du tableau 7.2). Le bénéfice immédiat tiré de cette méthode provient de la rapidité d'exécution de l'opération de *ET logique*, effectuée en moins d'une microseconde. La complexité de l'opération est en $\mathcal{O}(\mathrm{E}(n/w))$, avec n la longueur du mot binaire et w la taille des mots utilisés pour stocker l'opération logique. Le gain théorique atteint un facteur 50 par rapport aux méthodes d'intersection d'ensembles tels que les TreeSet. Il est important de noter qu'à aucun moment, le treillis, ou même une partie du treillis, n'a à être généré.

	o1	o2	o3	o4	o5
attribut_1	1	1	1	1	0
attribut_2	1	0	0	1	0
attribut_3	0	0	0	1	1
Extent	0	0	0	1	0

TABLEAU 7.2 – Une opération de ET logique permet de calculer l'extension d'un ensemble d'attributs

7.3.4 Application et évaluation

Nous avons testé notre implémentation sur un contexte extrait de la base eBDtheque (voir le chapitre 5 pour une description de son contenu). Les 849 cases de la base ont constitué l'ensemble des observations, tandis que leurs différents éléments de description ont composé l'ensemble des attributs. Ces éléments de description

sont extraits de notre ontologie et peuvent porter sur :

- les caractéristiques des cases, telles que leurs dimensions, leur forme ou leur emplacement dans la séquence de lecture ;
- leur contenu, comme par exemple la présence de bulles, de personnages, de certaines occurrences de mots, etc. ;
- leur contenant, principalement des informations bibliographiques sur la bande dessinée dont la case est tirée.

L'intégralité de ces éléments a été répartie en deux ensembles d'attributs, le premier contenant 100 éléments, le second 3533. Nous effectuons nos tests sur deux contextes distincts $K_1 = (O_1; I_1, (\alpha, \beta))$ et $K_2 = (O_2; I_2, (\alpha, \beta))$ avec $|O_1| = |O_2| = 849$, $|I_1| = 100$ et $|I_2| = 3533$.

Les intensions des éléments de O_1 sont en moyenne composées de 7 éléments, avec une borne minimum à 2 et une borne maximum à 11. Parmi les éléments de I_1, 28 ne font partie de l'intension que d'une seule observation. Le deuxième contexte présente un visage sensiblement différent. Les éléments de O_2 ont des intensions composées d'en moyenne 15 attributs. L'ensemble I_2, incluant par ailleurs l'ensemble I_1, contient 3403 éléments qui ont une extension de cardinalité inférieure ou égale à trois. Seulement 15 attributs ont, à l'inverse, une extension composée de plus de 100 éléments.

Le tableau 7.3 présente un court extrait du contexte, également illustré par la figure 7.9.

	panel 1	panel 2	panel 3	panel 4	panel 5
contains :balloon	1	1	1	1	0
shape :wide	0	1	0	0	1
shape :high	1	0	0	1	0
size :medium	0	0	1	0	1
contains :character	0	0	0	1	1

TABLEAU 7.3 – Contexte d'exemple

Si l'on veut qu'un tel système puisse être utilisé pour faire de la recherche d'informations, a fortiori dans un contexte d'interaction avec un utilisateur, il doit être efficace dans les deux sens. D'un concept vers ses successeurs mais également vers ses prédécesseurs. Nous avons donc testé notre implémentation sur le calcul des successeurs immédiats du concept absurde $\perp$ et sur les prédécesseurs immédiats du concept général $\top$. Les prédécesseurs immédiats d'un concept d'un contexte K sont vus comme les successeurs immédiats de ce même concept généré à partir d'un

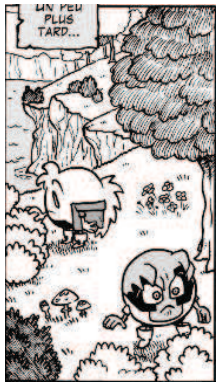


FIGURE 7.9 – Case correspondant à l'intension de la ligne 4 du tableau 7.3. Crédits : Cyb - Bubblegôm Gôm Tome 1 - Studio Cyborga.

contexte K^{-1}, inverse de K. Un contexte inversé est un contexte pour lequel les attributs ont pris la place des observations et vice versa. Les prédécesseurs immédiats de $\top_K$ sont les successeurs immédiats de $\bot_{K^{-1}}$. Les treillis des figures 7.10 et 7.11 permettent de se rendre compte du phénomène.

Le tableau 7.4 présente les temps de calcul mesurés pour une implémentation de l'algorithme de Bordat, une implémentation de LOA et une implémentation de LOA fonctionnant sur le principe des mots binaires. Les successeurs et les prédécesseurs ont été calculés pour les deux contextes K_1 et K_2 et leurs inverses. Les traitements ont été effectués sur une machine équipée d'un processeur à quatre cœurs cadencés à 2,7 GHz, épaulé par 8 Go de DDR3.

	Successeurs immédiats		Prédécesseurs immédiats	
	$\lvert O_1\rvert = 848$ $\lvert I_1\rvert = 100$	$\lvert O_1\rvert = 848$ $\lvert I_2\rvert = 3533$	$\lvert O_1\rvert = 100$ $\lvert I_1\rvert = 848$	$\lvert O_2\rvert = 3533$ $\lvert I_1\rvert = 848$
Classical + TreeSet	3.06	11767.52	549.76	994.00
Classical + BitSet	0.77	196.58	62.39	9.77
LOA + TreeSet	0.29	11.26	5.65	1183.75
LOA + BitSet	0.02	0.15	0.24	1.20

TABLEAU 7.4 – Temps de calcul (en secondes) des successeurs immédiats de $\bot$ et des prédécesseurs immédiats de $\top$.

Les résultats font état d'une diminution significative des temps de traitement pouvant être d'une part attribuée à LOA, d'autre part à l'utilisation de l'optimisation binaire. Cette optimisation, opérée sur l'algorithme *Limited Object Access*, procure en effet, selon les cas, un facteur de réduction computationnelle pouvant

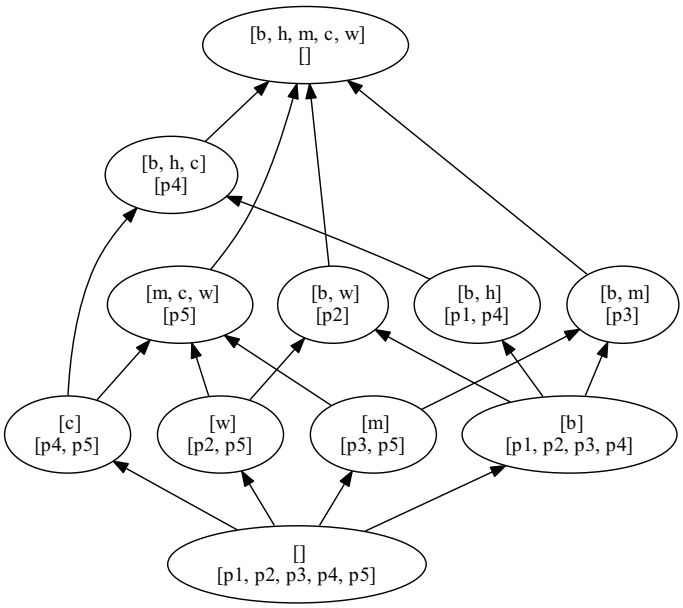


FIGURE 7.10 – Treillis des concepts généré à partir du contexte du tableau 7.3. Les attributs constituent l'intension des concepts, les observations forment leurs extensions.

aller de 14 à plus de 900. Le calcul des prédécesseurs immédiats de $\perp$ dans K_2 est le pire cas, avec 848 concepts différents comprenant chacun une observation (une case) et un ensemble d'attributs (sa description complète). Le temps de traitement est entièrement dévolu au calcul des extensions (plus d'un million d'appels), ce qui explique le gain spectaculaire apporté par les mots binaires dans ce cas de figure. Le gain de performances apporté par l'utilisation de mots binaires est moindre, mais significatif, sur l'algorithme de Bordat. Le temps alloué au calcul des extensions est en effet moins prédominant pour cet algorithme que pour LOA.

Une mesure supplémentaire a été effectuée sur les successeurs immédiats de 500 concepts sélectionnés au hasard dans le contexte K_1. Le temps de calcul moyen nécessaire à la génération des successeurs est de 0,18 seconde.

Nous obtenons toutefois des temps de calcul restant sous la barre de la seconde, sur une machine ordinaire, pour des jeux de données assez conséquents. Cela permet d'ouvrir des perspectives intéressantes pour la création de système de recherche interactive, où un bouclage de pertinence avec un utilisateur du système implique une réactivité sans faille, même dans le cas où de multiples calculs de succession doivent être effectués.

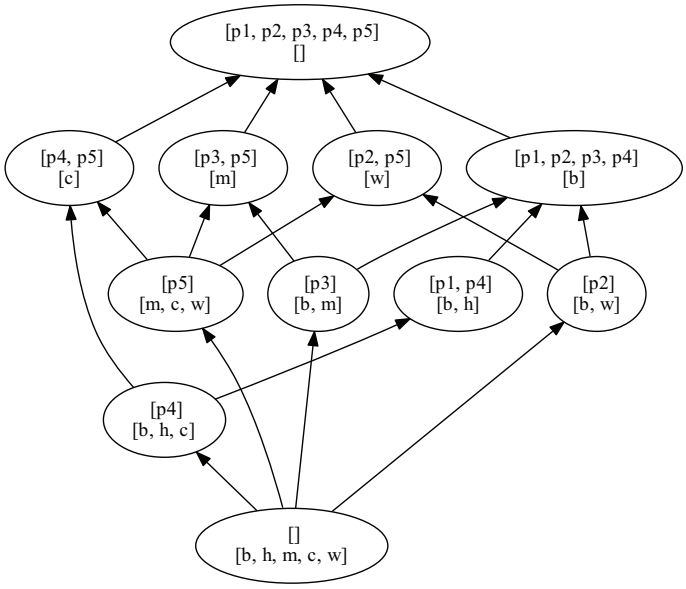


FIGURE 7.11 – Treillis des concepts généré à partir de l'inversion du contexte du tableau 7.3. Les observations constituent l'intension des concepts, les attributs forment leurs extensions.

Ces travaux ont fait l'objet d'une présentation lors de la dixième édition de CLA, à la Rochelle en 2013.

7.4 Conclusion

Nous avons présenté dans ce dernier chapitre comment notre conceptualisation peut être utilisée pour faciliter la navigation au sein d'une bande dessinée. Les méthodes présentées ne sont qu'une illustration de ce qu'il est possible de faire et peuvent être appliquées aux bulles et au texte, ou encore étendues d'une planche à l'autre. D'une manière générale, notre ontologie a été conçue pour permettre l'inférence d'informations sur les éléments plutôt que l'inférence de connaissances sur les concepts la composant. Sur ce point, quelques modifications apportées à l'ontologie de bande dessinée permettront de faire émerger de nouvelles capacités de déduction. Un relâchement des contraintes sur les propriétés liant les concepts de contenu (*Panel*, *TextLine*, etc.) et l'ajout de définitions à ces concepts permettra une classification plus fine de leurs instances. Ces définitions sont dépendantes du contexte d'application et doivent être établies conformément à l'objectif d'utilisation final.

Ces nouvelles informations peuvent s'inscrire dans une base de connaissances au sein de laquelle des recherches d'informations complexes pourront être effectuées. Les résultats obtenus pour le parcours des concepts d'un treillis ouvrent en effet des perspectives intéressantes concernant la recherche et l'annotation interactive d'une telle base.

Chapitre 8

Conclusion générale

Sommaire

8.1 Conclusion 149
8.1.1 Deux ontologies et une base annotée... 150
8.1.2 ...pour la compréhension d'images... 151
8.1.3 ...et la recherche d'informations dans une BD. 151
8.2 Perspectives 151
8.2.1 Enrichir les interactions,... 151
8.2.2 ...les capacités de raisonnement... 152
8.2.3 ...et la recherche d'informations. 152

8.1 Conclusion

Nous touchons à la fin de la présentation de ces travaux de thèse, il est maintenant temps d'en dresser un bilan et de conclure. Cette thèse s'est inscrite dans le cadre du projet eBDtheque comprenant les travaux de deux doctorants et de leurs encadrants, dont ceux présentés dans ce manuscrit, épaulés par un ingénieur et un post-doc au cours de la première année. Les objectifs poursuivis lors de la mise en place de ce projet portaient sur la fourniture d'outils et de techniques permettant d'apporter une réponse aux besoins des acteurs socio-économiques du monde de la bande dessinée numérique. Ces besoins s'articulent notamment autour de la notion de compréhension du contenu des planches, d'un point de vue visuel et sémantique. L'extraction et l'analyse sémantique automatique de ce contenu permet de produire un cadre à partir duquel de multiples applications peuvent être envisagées. Celles-ci incluent notamment la transformation assistée du contenu, de manière pertinente, vers de nouveaux supports numériques ou encore la recherche de scènes ou d'éléments complexes au sein de larges bases d'albums numérisés et annotés.

La partie de ces travaux correspondant au traitement et à l'extraction du contenu des images numérisées a été réalisée dans le cadre de la deuxième thèse de ce projet, menée en parallèle au sein du L3i et du CVC par Christophe Rigaud [Rigaud 2014a]. Les travaux présentés dans ce manuscrit avaient pour objectif de poser un cadre structurant le contenu extrait. Ce cadre devait, dans un premier temps, permettre l'analyse sémantique et la déduction d'informations à partir des éléments présents dans les planches et de leurs interactions. Dans un second temps, une synergie devait

être mise en place entre le système d'extraction des caractéristiques visuelles et le modèle de raisonnement.

8.1.1 Deux ontologies et une base annotée...

Nous avons développé et détaillé dans ce manuscrit une conceptualisation permettant d'atteindre ces buts. Elle est basée sur deux ontologies distinctes, émanant de l'étude du domaine de l'analyse d'images et du domaine de la bande dessinée. Ces deux ontologies, lorsqu'elles entrent en conjonction, permettent d'une part de manipuler les données extraites comme des éléments de bandes dessinées, obéissant de fait aux codes et aux règles tacites de l'art, et, d'autre part, de manipuler le contenu d'une bande dessinée comme de simples zones dans l'image pouvant être manipulées, transformées et mises en relation les unes avec les autres.

L'ontologie image présentée ne redéfinit pas de standards mais en assure la compatibilité dans ses choix de conception. Elle présente l'avantage d'être simple, de ne comporter qu'un petit nombre de concepts et d'être indépendante d'un domaine d'application particulier, tout en restant efficace pour les tâches que nous nous sommes fixées. Elle peut facilement s'exporter dans d'autres domaine que celui de la bande dessinée, notamment au sein du laboratoire.

L'ontologie conceptualisant le domaine de la bande dessinée ayant été présentée a été pensée pour être utilisée comme un support pour l'analyse des images. Comme cela a été souligné, la bande dessinée est un domaine possédant des codes et un vocabulaire qui se retrouvent, majoritairement, dans un volume important d'œuvres. Ces codes peuvent être formalisés et un certain nombre de choix de conceptualisation effectués se retrouvent notamment dans l'ontologie développée en parallèle par [Rissen 2012]. L'art en lui-même se doit cependant, par définition, de briser ces codes, d'expérimenter pour aboutir à de nouvelles choses. Le patrimoine mondial de la bande dessinée présente en conséquence un visage très hétérogène. La composition des éléments d'une BD ne suivant pas de règle formelle, la seule propriété véritablement transversale au domaine est la notion de « dessins juxtaposés en séquence ». Dès lors, la proposition d'une conceptualisation valide dans sa représentation de toutes les instances de bandes dessinées serait trop souple dans ses contraintes pour l'analyse d'images à laquelle elle est destinée.

La pertinence des choix effectués a tout de même pu être évaluée sur une base de données annotées, concluant à une capacité de représentation d'environ 90% des éléments pris en charge (cases, bulles, texte et personnages). La conception et la construction de cette base ont par ailleurs été des contributions importantes apportées à la communauté scientifique. C'est, à ce jour et à notre connaissance, la seule base proposant une vérité terrain pour l'analyse visuelle et sémantique de la bande dessinée. Elle a été rendue publique en 2013 et téléchargée par des équipes provenant de divers endroits du globe.

8.1.2 ...pour la compréhension d'images...

Les deux ontologies développées ont été mises en œuvre au sein d'un framework d'analyse d'images de bandes dessinées. A travers un mécanisme de bouclage entre les algorithmes d'extraction et le système de représentation des connaissances et de raisonnement, les données extraites sont organisées au sein d'une structure formelle. L'information déduite lors de ces interactions permet d'accroître de manière itérative la compréhension globale du contenu d'une page, de guider l'extraction de données plus complexes et d'augmenter significativement la précision des résultats.

8.1.3 ...et la recherche d'informations dans une BD.

D'un autre côté, les ontologies proposées peuvent également être mises au service de la recherche d'informations et d'un enrichissement sémantique plus profond des données. Le mécanisme de raisonnement des ontologies est basé sur une logique de monde ouvert. C'est une force lorsque l'on traite des données incomplètes ou incertaines, comme c'est le cas lors de la phase d'analyse d'images, mais cela peut parfois poser problème lorsque l'on traite, au contraire, des données complètes et sûres. Les règles de déduction doivent alors être adaptées en tenant compte de cette propriété, dans un contexte où l'absence d'objet peut parfois signifier beaucoup. La recherche d'informations au sein de bases annotées a été abordée et des solutions techniques ont été proposées pour faciliter l'interaction entre le système et l'utilisateur.

Enfin, la propriété principale de la bande dessinée qui est le séquencement des images véhiculant le propos de l'auteur a reçu une attention particulière. L'ordonnancement des instants de l'histoire a été mis en place, tirant partie des propriétés des ontologies pour intégrer la souplesse de lecture que l'on peut trouver dans le parcours d'une planche classique de bande dessinée.

8.2 Perspectives

Les objectifs initiaux du projet eBDtheque étaient ambitieux et plusieurs points auxquels nous avons pensé n'ont pas pu être intégrés dans le cadre de cette thèse. Ils font néanmoins l'objet de perspectives à plus ou moins long terme portant sur l'évolution du cadre pour, d'un côté, enrichir les interactions et le raisonnement et, d'un autre côté, améliorer la recherche d'informations.

8.2.1 Enrichir les interactions,...

Tout d'abord, le système d'analyse globale des planches peut être enrichi, à court terme, de différentes façons. Il est notamment possible de multiplier les interactions entre les aspects bas niveau et haut niveau afin de corriger plus efficacement les extractions. La présence de phylactères vides de toute ligne de texte pourrait par

exemple donner lieu à un retour vers les extracteurs qui, en ajustant leurs paramètres au nouveau contexte d'image, chercheraient la présence de lignes ayant été manquées initialement. Réciproquement, un groupement spatial de lignes de texte au sein d'une case pourrait déclencher la recherche d'une bulle manquée entourant ces lignes.

8.2.2 ...les capacités de raisonnement...

D'une manière générale, une intégration plus profonde des raisonnements spatiaux, à travers l'apport d'un standard tel que GeoSPARQL dans l'ontologie image, permettrait l'intégration de règles logiques basées sur le positionnement spatial des éléments au sein des cases. Cela rendrait par exemple possible la différentiation des cadres de narration, dont la position est traditionnellement proche du coin d'une case, des faux positifs de bulle ne possédant pas de flèche. De même, les techniques de raisonnement spatial basées sur la logique floue constituent une piste intéressante pour l'amélioration de la détection des personnages à partir de la direction des flèches. C'est l'aspect de notre système qui possède la plus grande marge de progression.

De plus, l'extension de notre conceptualisation du domaine de la bande dessinée est également envisagée à plusieurs niveaux. L'intégration d'un concept représentant un groupe de cases, représentant chacune un plan d'une scène ou d'une séquence particulière, serait notamment intéressante.Cela permettrait d'intégrer un niveau intermédiaire entre la planche et la case, correspondant potentiellement au *strip* du comics américain, permettant une meta-analyse d'une scène, transversale à plusieurs cases. L'intégration d'ontologies pour l'annotation du contenu, telles que celles basées sur des extractions de WordNet utilisées dans les travaux de [Zinger 2005], pourrait également se révéler pertinente à moyen terme. L'annotation du contenu des scènes représentées dans les cases, au delà des personnages, augmenterait significativement le potentiel de la recherche d'informations.

8.2.3 ...et la recherche d'informations.

Enfin, les aspects portant sur la recherche d'informations au sein de ces bases de connaissances vont être approfondis. Les contributions présentées vont dans le sens d'un système basé sur une interaction avec l'utilisateur, tenant compte de ses actions afin de le guider efficacement dans sa recherche ou dans son annotation. Les approches développées dans les travaux de [Ferré 2002] sont source d'inspiration et les moyens de transposer la structure d'une base de connaissance dans une structure de treillis sont étudiés.

Publications

Journal

Knowledge-driven understanding of images in comic books
Christophe Rigaud, Clément Guérin, Dimosthenis Karatzas, Jean-Christophe Burie et Jean-Marc Ogier
Submitted to the International Journal on Document Analysis and Recognition (IJDAR), Accepted on November 2014

Conférences

Réduction de l'espace de recherche pour les personnages de bandes dessinées
Clément Guérin, Christophe Rigaud, Karell Bertet, Jean-Christophe Burie, Arnaud Revel et Jean-Marc Ogier
Actes du 19^{me} Congrès National sur la Reconnaissance de Formes et l'Intelligence Artificielle (RFIA), Rouen, 2014

An efficient Java implementation of the immediate successors calculation
Clément Guérin, Karell Bertet et Arnaud Revel
Proceedings of the 10^{th} International Conference on Concept Lattices and Their Applications (CLA), La Rochelle, pp. 81-92, 2013

eBDtheque: a representative database of comics
Clément Guérin, Christophe Rigaud, Antoine Mercier, Farid Ammar-Boudjelal, Karell Bertet, Alain Bouju, Jean-Christophe Burie, Georges Louis, Jean-Marc Ogier et Arnaud Revel
Proceedings of the 12^{th} International Conference on Document Analysis and Recognition (ICDAR), Washington DC, pp. 1145-1149, 2013

Ontologies et relations spatiales dans la lecture d'une bande dessinée
Norbert Tsopze, Clément Guérin, Karell Bertet et Arnaud Revel
Actes des 23^{me} Journées Francophones d'Ingénierie des Connaissances (IC), Paris, pp. 175-182, 2012

Workshops

Localisation contextuelle des personnages de bandes dessinées
Christophe Rigaud et Clément Guérin
Actes du 8^{me} Colloque International Francophone sur l'Ecrit et le Document (CIFED), Nancy, 2014

Ontologies and spatial relations applied to comicbooks reading
Clément Guérin
Presented at the Phd. Symposium of the 18^{th} International Conference on Knowledge Engineering and Knowledge Management (EKAW), Galway, 2012

An approach to Sematic Content Based Image Retrieval using Logical Concept Analysis. Application to comicbooks.
Clément Guérin, Karell Bertet et Arnaud Revel
Presented at "What can FCA do for AI ?", Workshop of the 20^{th} European Conference on Artificial Intelligence (ECAI), Montpellier, 2012

Bibliographie

[Allen 1996] R. B. Allen. *Retrieval from facet spaces.* Eletronic Publishing, vol. 8, no. 2-3, pages 247–257, 1996. (Cité en page 135.)

[Amigoni 2005] Francesco Amigoni et Mario Arrigoni Neri. *An Application of Ontology Technologies to Robotic Agents.* In Intelligent Agent Technology, pages 751–754, 2005. (Cité en page 34.)

[Anand 2010] Sanjay Anand et Akshat Verma. *Development of Ontology for Smart Hospital and Implementation using UML and RDF.* International Journal of Computer Science Issues, vol. 7, no. 5, pages 206–212, 2010. (Cité en page 33.)

[Arai 2010] Kohei Arai et Herman Tolle. *Method for Automatic E-Comic Scene Frame Extraction for Reading Comic on Mobile Devices.* In Seventh International Conference on Information Technology : New Generations, pages 370–375, 2010. (Cité en page 47.)

[Arai 2011] Kohei Arai et Herman Tolle. *Method for Real Time Text Extraction of Digital Manga Comic.* International Journal of Image Processing (IJIP), vol. 4, no. 6, pages 669–676, 2011. (Cité en page 47.)

[Auer 2007] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak et Zachary Ives. *DBpedia : A Nucleus for a Web of Open Data.* In International Semantic Web Conference, 2007. (Cité en page 17.)

[Baader 1996] Franz Baader. *A Formal Definition for the Expressive Power of Terminological Knowledge Representation Languages.* Journal of Logic and Computation, vol. 6, no. 1, pages 33–54, 1996. (Cité en page 30.)

[Baader 2003] Franz Baader, Deborah McGuinness, Daniele Nardi et Peter Patel-Schneider. The Description Logic Handbook–Theory, Implementation and Applications. 2003. (Cité en pages 28 et 32.)

[Bada 2004] Michael Bada, Robert Stevens, Carole Goble, Yolanda Gil, Michael Ashburner, Judith A. Blake, J. Michael Cherry, Midori Harris et Suzanna Lewis. *A Short Study on the Success of the Gene Ontology.* Web Semantics : science, services and agents on the World Wide Web, vol. 1, no. 2, pages 235–240, 2004. (Cité en page 33.)

[Bannour 2011] Hichem Bannour et Céline Hudelot. *Towards ontologies for image interpretation and annotation.* In Content-Based Multimedia Indexing (CBMI), pages 211–216. IEEE, 2011. (Cité en pages 34 et 56.)

[Battle 2012] Robert Battle et Dave Kolas. *Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL.* Semantic Web, vol. 3, no. 4, pages 355–370, 2012. (Cité en page 60.)

[Belgiu 2014] Mariana Belgiu, Barbara Hofer et Peter Hofmann. *Coupling formalized knowledge bases with object-based image analysis.* Remote Sensing Letters, vol. 5, no. 6, pages 530–538, Juillet 2014. (Cité en page 59.)

[Benhamou 2011] Françoise Benhamou, Nathalie Moureau, Stéphanie Peltier et Olivia Guillon. *La filière du livre.* In Culture numérique : regards sur les industries créatives - ANR PANIC, pages 27–60. Le Manuscrit, 2011. (Cité en page 5.)

[Bergaglio 2004] Sylvie Bergaglio et Claude Tournet. *Etude des codes de la BD à travers Maus*, 2004. (Cité en pages 36 et 45.)

[Berners-Lee 1991] Tim Berners-Lee, Robert Cailliau, Jean-François Groff et Bernd Pollermann. *World-Wide Web : The Information Universe.* Internet Research, vol. 2, no. 1, pages 52–58, 1991. (Cité en page 14.)

[Berners-Lee 2001] Tim Berners-Lee, James Hendler et Ora Lassila. *The Semantic Web.* vol. 21, 2001. (Cité en page 16.)

[Beylie 1964] Claude Beylie. *La Bande dessinée est-elle un art ?* Lettres et Médecins, 1964. (Cité en page 1.)

[Birkhoff 1967] Garrett Birkhoff. Lattice Theory. 1967. (Cité en page 135.)

[Bizer 2009] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak et Sebastian Hellmann. *DBpedia - A crystallization point for the Web of Data.* Web Semantics : Science, Services and Agents on the World Wide Web, vol. 7, no. 3, pages 154–165, 2009. (Cité en page 17.)

[Blaschke 2010] T. Blaschke. *Object based image analysis for remote sensing.* ISPRS Journal of Photogrammetry and Remote Sensing, vol. 65, no. 1, pages 2–16, 2010. (Cité en page 59.)

[Bobillo 2011] Fernando Bobillo et Umberto Straccia. *Fuzzy Ontology Representation using OWL 2.* International Journal of Approximate Reasoning, vol. 52, no. 7, pages 1073–1094, 2011. (Cité en page 59.)

[Bordat 1986] J.-P. Bordat. *Calcul pratique du treillis de Galois d'une correspondance.* Math. Sci. Hum., vol. 96, pages 31–47, 1986. (Cité en pages 137 et 138.)

[Borst 1997] Willem Nico Borst. *Construction of engineering ontologies for knowledge sharing and reuse.* PhD thesis, University of Twente, 1997. (Cité en page 20.)

[Bouissou 2008] J.-M. Bouissou. *Pourquoi le manga est-il devenu un produit culturel global ?* Esprit, no. 346, 2008. (Cité en page 4.)

[Bournez 2013] Olivier Bournez. *Calcul propositionnel. Calcul des prédicats. Complétude*, 2013. (Cité en page 26.)

[Brachman 1984] Ronald J. Brachman et Hector J. Levesque. *The Tractability of Subsumption in Frame-Based Description Languages.* In AAAI, pages 34–37, 1984. (Cité en page 30.)

[Brachman 1985] Ronald J. Brachman et James G. Schmolze. *An Overview of the KL-ONE Knowledge Representation System.* Cognitive Science, vol. 9, no. 2, pages 171–216, 1985. (Cité en page 23.)

[Bray 1998] Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau et John Cowan. *Extensible Markup Language (XML)*, 1998. (Cité en page 17.)

[Brickley 1999] Dan Brickley et R. V. Guha. *Resource description framework (RDF) model and syntax specification*. Rapport technique, 1999. (Cité en page 17.)

[Brickley 2000] Dan Brickley et R. V. Guha. *Resource Description Framework (RDF) Schema Specification 1.0 : W3C Candidate Recommendation 27 March 2000*. Rapport technique, W3C, 2000. (Cité en page 18.)

[Brickley 2006] Dan Brickley. *Basic Geo (WGS84 lat/long) Vocabulary*, 2006. (Cité en page 52.)

[Butler 2008] Howard Butler, Martin Daly, Allan Doyle, Sean Gillies, Tim Schaub et Christopher Schmidt. *The GeoJSON Format Specification*. Rapport technique, 2008. (Cité en page 67.)

[Canudo 1923] Ricciotto Canudo. *Manifeste des Sept Arts*. Gazette des sept arts, vol. 2, 1923. (Cité en page 1.)

[Carpineto 1995] Claudio Carpineto et Giovanni Romano. *ULYSSES : A Lattice-based Multiple Interaction Strategy Retrieval Interface*. In East-West Human-Computer Interaction Conference (EWHCI), pages 91–104, 1995. (Cité en page 136.)

[Caspard 2003] Nathalie Caspard et Bernard Monjardet. *The lattices of closure systems, closure operators, and implicational systems on a finite set : a survey*. Discrete Applied Mathematics, vol. 127, no. 2, pages 241–269, 2003. (Cité en page 138.)

[CCI 2008] CCI. *Le marché de la vidéo à la demande (VOD) en France : un aperçu des pratiques, des enjeux et des perspectives de développement*. Rapport technique, 2008. (Cité en page 5.)

[Chang 1998] Shih-Fu Chang, William Chen et Hari Sundaram. *Semantic Visual Templates : Linking Visual Features to Semantics*. In International Conference on Image Processing, pages 531–535, 1998. (Cité en page 56.)

[Chella 2002] Antonio Chella, Massimo Cossentino, Roberto Pirrone et Andrea Ruisi. *Modeling Ontologies for Robotic Environments*. In International Conference on Software Engineering and Knowledge Engineering, pages 77–80, 2002. (Cité en page 34.)

[Chen 2012] Na Chen, Qian-Yi Zhou et Viktor K. Prasanna. *Understanding web images by object relation network*. International conference on World Wide Web, pages 291–300, 2012. (Cité en page 60.)

[Chiron 2008] Julia Chiron. *Le neuvième art sur Internet : les blogs de bande dessinée*. PhD thesis, 2008. (Cité en page 5.)

[Christodoulou 2012] George Christodoulou, Euripides G.M. Petrakis et Sotirios Batsakis. *Qualitative Spatial Reasoning Using Topological and Directional*

Information in OWL. In International Conference on Tools with Artificial Intelligence, pages 596–602. Ieee, 2012. (Cité en pages 58 et 60.)

[Chung Ho 2007] Chan Chung Ho, Leung Howard et Komura Taku. *Automatic panel extraction of color comic images.* Advances in Multimedia Information Processing, vol. 4810, pages 775–784, 2007. (Cité en page 46.)

[CIBDI 2012] CIBDI. *Archives Excelsior Dimanche - Dimanche illustré*, 2012. (Cité en page 41.)

[CNL 2010] CNL. *Le livre sera-t-il numérique ?* Rapport technique, 2010. (Cité en page 5.)

[Cohn 1997] Anthony G. Cohn, Brandon Bennett, John Gooday et Nicholas Mark Gotts. *Representing and reasoning with qualitative spatial relations about regions.* 1997. (Cité en pages 60 et 127.)

[Collectif PANIC 2011] Collectif PANIC. Culture Numérique. Regards sur les industries créatives. Le Manuscrit, 2011. (Cité en page 6.)

[Coogan 2009] Peter Coogan. *The definition of the superhero.* In A comics studies reader, pages 77–93. 2009. (Cité en page 3.)

[Cook 1971] Stephen A. Cook. *The complexity of theorem-proving procedures.* Symposium on Theory Of Computing, pages 151–158, 1971. (Cité en page 24.)

[Coustaty 2011] Mickael Coustaty, Alain Bouju, Karell Bertet et Georges Louis. *Using ontologies to reduce the semantic gap between historians and image processing algorithms.* In International Conference on Document Analysis and Recognition (ICDAR), pages 156–160. IEEE, 2011. (Cité en page 58.)

[Crevier 1997] Daniel Crevier et Richard Lepage. *Knowledge-Based Image Understanding Systems : A Survey.* Computer Vision and Image Understanding, vol. 67, no. 2, pages 161–185, 1997. (Cité en page 52.)

[Cuenca Grau 2008] Bernardo Cuenca Grau, Ian Horrocks, Boris Motik, Bijan Parsia, Peter Patel-schneider et Ulrike Sattler. *OWL 2 : The Next Step for OWL.* Web Semantics : science, services and agents on the World Wide Web, vol. 6, no. 4, pages 309–322, 2008. (Cité en page 22.)

[D'Agostino 1999] Marcello D'Agostino. *Tableau Methods for Classical Propositionnal Logic.* In Handbook of Tableau Methods, pages 45–124. 1999. (Cité en page 27.)

[Date 1987] C.J. Date et Hugh Darwen. A Guide to the SQL Standard. Addison-Wesley New York, 1987. (Cité en page 19.)

[Decker 1983] Dwight Decker et Gary Groth. *An interview with William M. Gaines.* The Comics Journal, vol. 81, 1983. (Cité en page 3.)

[Demko 2011] Christophe Demko et Karell Bertet. *Generation algorithm of a concept lattice with limited object access.* In International Conference on Concept Lattices and Their Applications (CLA), pages 239–250, 2011. (Cité en pages 139 et 140.)

[Deyzieux 2008] Agnès Deyzieux. *Les grands courants de la bande dessinée.* Le français aujourd'hui, vol. 161, pages 59–68, 2008. (Cité en page 2.)

[Dogan 2013] Belya Dogan. *BD marché mondial, particularités du style Franco-belge en BD*, 2013. (Cité en page 2.)

[Donini 1997] Francesco M. Donini, Maurizio Lenzerini et Daniele Nardi. *The Complexity of Concept Languages.* Information and Computation, vol. 134, no. 1, pages 1–58, 1997. (Cité en page 30.)

[Dousset 2004] Claire Dousset. *Approche de la bande dessinée grâce aux onomatopées : de l'étude à la production.* Rapport technique, 2004. (Cité en page 45.)

[Duc 1982] Bernard Duc. L'art de la BD - Tome 1 - Du scénario à la réalisation. Glénat édition, 1982. (Cité en page 45.)

[Duc 1983] Bernard Duc. L'art de la BD - Tome 2 - La technique du dessin. Glénat édition, 1983. (Cité en page 38.)

[Duda 1972] Richard O. Duda et Peter E. Hart. *Use of the hough transformation to detect lines and curves in pictures.* In ACM international conference on Multimedia, volume 15, pages 11–15, 1972. (Cité en page 46.)

[Eisner 1997] Will Eisner. La Bande dessinée, art séquentiel. Vertige gr édition, 1997. (Cité en page 41.)

[Eisner 1998] Will Eisner. *Le récit graphique.* Paris, Vertige Graphic, 1998. (Cité en page 36.)

[Everingham 2005] Mark Everingham, Andrew Zisserman, Christopher K. I. Williams et Luc Gool. *The 2005 PASCAL Visual Object Classes challenge.* In First PASCAL Challenge Workshop, 2005. (Cité en page 54.)

[Everingham 2009] Mark Everingham, Luc Gool, Christopher K. I. Williams, John Winn et Andrew Zisserman. *The Pascal Visual Object Classes (VOC) Challenge.* International Journal of Computer Vision, vol. 88, no. 2, pages 303–338, 2009. (Cité en pages 75, 88 et 118.)

[Ferraiolo 2001] Jon Ferraiolo. *Scalable Vector Graphics (SVG) 1.0 Specification.* Rapport technique, 2001. (Cité en page 67.)

[Ferré 2002] Sébastien Ferré et Olivier Ridoux. *Introduction to Logical Information Systems.* Rapport technique 3, 2002. (Cité en pages 136 et 152.)

[Ferré 2011] Sébastien Ferré et Alice Hermann. *Semantic search : reconciling expressive querying and exploratory search.* In International Semantic Web Conference, pages 177–192, 2011. (Cité en page 136.)

[Fischer 1971] M.J. Fischer et A.R. Meyer. *Boolean matrix multiplication and transitive closure.* In 12th Annual Symposium on Switching and Automata Theory, pages 129–131, 1971. (Cité en page 138.)

[Fitting 1996] Melvin Fitting. First-order logic and automated theorem proving. Springer, 1996. (Cité en page 27.)

[Forestier 2012] G. Forestier, A. Puissant, C. Wemmert et P. Gançarski. *Knowledge-based region labeling for remote sensing image interpretation.* Computers, Environment and Urban Systems, vol. 36, no. 5, pages 470–480, 2012. (Cité en page 61.)

[Forestier 2013] Germain Forestier, C. Wemmert et A. Puissant. *Coastal Image Interpretation Using Background Knowledge and Semantics.* Computers & Geosciences, vol. 54, no. 4, pages 88–96, 2013. (Cité en page 61.)

[François 2005] Virginie François. La bande dessinée. Scala édition, 2005. (Cité en page 2.)

[Gabilliet 2005] Jean-Paul Gabilliet. *Fun in four colors.* Transatlantica, vol. 1, 2005. (Cité en page 4.)

[Gabilliet 2009] Jean-Paul Gabilliet. *BD, mangas et comics : différences et influences.* Hermès, La Revue, vol. 54, no. 2, pages 35–40, 2009. (Cité en page 4.)

[Ganter 1984] Bernhard Ganter. *Two basic algorithms in concept analysis.* Technische Hochschule Darmstadt (Preprint 831), pages 312–340, 1984. (Cité en page 137.)

[Ganter 1999] Bernhard Ganter et Rudolf Wille. Formal Concept Analysis. Springer, 1999. (Cité en page 137.)

[Garrity 2011] Shaenon Garrity. *The History of Webcomics.* The Comics Journal, 2011. (Cité en page 5.)

[Gasevic 2009] Dragan Gasevic, Nima Kaviani et Milan Milanovic. *Ontologies and Software Engineering.* In Handbook on Ontologies, pages 593–616. 2009. (Cité en page 34.)

[Generator Research 2013] Generator Research. *Digital Music Subscription Services : 2013.* Rapport technique, 2013. (Cité en page 5.)

[Giles 2005] Jim Giles. *Internet encyclopaedias go head to head.* Nature, vol. 438, no. 7070, pages 900–1, 2005. (Cité en page 15.)

[Godin 1989] Robert Godin, Jan Gecsei et Claude Pichet. *Design of a browsing interface for information retrieval.* In International Conference on Research and Development in Information Retrieval, pages 32–39, 1989. (Cité en page 136.)

[Godin 1993] Robert Godin, Rokia Missaoui et Alain April. *Experimental comparison of navigation in a Galois lattice with conventional information retrieval methods.* International Journal of Man-Machine Studies, vol. 38, no. 5, pages 747–767, 1993. (Cité en page 135.)

[Groensteen 1996] Thierry Groensteen. *Neuvième art : petite histoire d'une appellation non contrôlée.* Neuvième Art, vol. 1, page 4, 1996. (Cité en page 1.)

[Gruber 1993] Thomas R. Gruber. *A Translation Approach to Portable Ontology Specifications.* Knowledge Acquisition, vol. 5, no. 2, pages 199–220, 1993. (Cité en page 20.)

[Guarino 2009] Nicolas Guarino, Daniel Oberle et Steffen Staab. *What Is an Ontology ?* In Handbook on Ontologies, pages 1–18. 2009. (Cité en page 20.)

[Gudivada 1995] V. N. Gudivada et V. V. Raghavan. *Content-Based Image Retrieval Systems.* Computer, vol. 28, no. 9, pages 18–22, 1995. (Cité en page 56.)

[Guérin 2013] Clément Guérin, Christophe Rigaud, Antoine Mercier, Farid Ammar-Boudjelal, Karell Bertet, Alain Bouju, Jean-Christophe Burie, Georges Louis, Jean-Marc Ogier et Arnaud Revel. *eBDtheque : A Representative Database of Comics.* In 2013 12th International Conference on Document Analysis and Recognition, pages 1145–1149. IEEE, 2013. (Cité en page 86.)

[Guilbert 2011] Xavier Guilbert. *La légitimation en devenir de la bande dessinée*, 2011. (Cité en page 1.)

[Han 2007] Eunjung Han, Jonyeol Yang, Hwangkyu Yang et Keechul Jung. *Automatic Mobile Content Conversion Using Semantic Image Analysis.* LNCS, pages 298–307, 2007. (Cité en page 61.)

[Hanbury 2008] Allan Hanbury. *A Survey of Methods for Image Annotation.* Journal of Visual Languages And Computing, vol. 19, no. 5, pages 617–627, 2008. (Cité en page 54.)

[Harvey 1994] Robert C Harvey. The art of the funnies : an aesthetic history. 1994. (Cité en page 3.)

[Hay 2008] Geoffrey J. Hay et G. Castilla. *Geographic Object-Based Image Analysis (GEOBIA) : A new name for a new discipline.* In Thomas Blaschke, Stefan Lang et Geoffrey J. Hay, editeurs, Object-Based Image Analysis, pages 75–89. Springer Berlin Heidelberg, lecture no édition, 2008. (Cité en page 49.)

[Hermann 2012] Alice Hermann, Sébastien Ferré et Mireille Ducassé. *An Interactive Guidance Process Supporting Consistent Updates of RDFS Graphs.* In International Conference on Knowledge Engineering and Knowledge Management, pages 185–199, 2012. (Cité en page 136.)

[Hitzler 2009] Pascal Hitzler, Markus Krötzsch et Sebastian Rudolph. *Knowledge Representation for the Semantic Web Part II : Rules for OWL*, 2009. (Cité en page 22.)

[Ho 2011] Anh Khoi Ngo Ho, Jean-Christophe Burie et Jean-Marc Ogier. *Comics page structure analysis based on automatic panel extraction.* In 9th IAPR International Workshop on Graphics Recognition, 2011. (Cité en page 47.)

[Ho 2012] Anh Khoi Ngo Ho, Jean-Christophe Burie et Jean-Marc Ogier. *Panel and Speech Balloon Extraction from Comic Books.* In 10th IAPR International Workshop on Document Analysis Systems, pages 424–428. Ieee, 2012. (Cité en page 47.)

[Ho 2013] Hoang Nam Ho, Christophe Rigaud, Jean-Christophe Burie et Jean-Marc Ogier. *Redundant structure detection in attributed adjacency graphs for character detection in comics books.* In International Workshop on Graphics Recognition (GREC), 2013. (Cité en page 48.)

[Hollink 2003] Laura Hollink, A. Th. (Guus) Schreiber, Jan Wielemaker et Bob Wielinga. *Semantic Annotation of Image Collections.* In KCAP'03 Workshop on Knowledge Capture and Semantic Annotation, volume 2, 2003. (Cité en page 54.)

[Horridge 2006] Matthew Horridge, Nick Drummond, John Goodwin, Alan Rector, Robert Stevens et Hai H. Wang. *The Manchester OWL Syntax.* In OWL : Experiences and Directions (OWLED), 2006. (Cité en page 73.)

[Horrocks 2003] Ian Horrocks et Peter F. Patel-Schneider. *Reducing OWL Entailment to Description Logic Satisfiability.* Semantic Web, vol. 2870, pages 17–29, 2003. (Cité en page 30.)

[Horrocks 2004] Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof et Mike Dean. *SWRL : A Semantic Web Rule Language Combining OWL and RuleML.* Rapport technique, 2004. (Cité en page 22.)

[Hu 2003] Bo Hu, S. Dasmahapatra, P. Lewis et N. Shadbolt. *Ontology-based medical image annotation with description logics.* In International Conference on Tools with Artificial Intelligence, pages 77–82, 2003. (Cité en pages 49 et 55.)

[Huang 2011] Shan Shan Huang, Todd J. Green et Boon Thau Loo. *Datalog and Emerging Applications : An Interactive Tutorial.* In International Conference on Managemente of Data, pages 1213–1216, 2011. (Cité en page 22.)

[Hudelot 2008a] Céline Hudelot. *Ontologies pour l'analyse et l'interprétation d'images : état de l'art.* In Journées Francophones Extraction et Gestion des Connaissances, pages 5–6, 2008. (Cité en page 52.)

[Hudelot 2008b] Céline Hudelot, Jamal Atif et Isabelle Bloch. *Fuzzy spatial relation ontology for image interpretation.* Fuzzy Sets and Systems, vol. 159, no. 15, pages 1929–1951, 2008. (Cité en pages ix, 58 et 59.)

[Hyvonen 2002] Eero Hyvonen, Avril Styrman et Samppa Saarela. *Ontology-based image retrieval.* Towards the semantic web and web, 2002. (Cité en page 54.)

[Jewell 2005] Michael O. Jewell, Faith Lawrence et Mischa M. Tufeld. *OntoMedia : An Ontology for the Representation of Heterogeneous Media.* In SIGIR Workshop on Multimedia Information Retrieval, 2005. (Cité en page 51.)

[Jiang 2004] Shuqiang Jiang, Tiejun Huang et Wen Gao. *An Ontology-based Approach to Retrieve Digitized Art Images.* In IEEE/WIC/ACM International Conference on Web Intelligence, pages 131–137, 2004. (Cité en page 58.)

[Johnson-Laird 1984] Philip N. Johnson-Laird, Douglas J. Herrmann et Roger Chaffin. *Only connections : A critique of semantic networks.* Psychological Bulletin, vol. 96, no. 2, pages 292–315, 1984. (Cité en page 23.)

[Kass 1988] Michael Kass, Andrew Witkin et Demetri Terzopoulos. *Snakes : Active Contour Models.* International Journal of Computer Vision, vol. 1, no. 4, pages 321–331, 1988. (Cité en page 47.)

[Kataria 2008] Pavandeep Kataria, Radmila Juric, Shamimabi Paurobally et Kambiz Madani. *Implementation of Ontology for Intelligent Hospital Wards.* In Hawaii Conference on System Sciences, 2008. (Cité en page 33.)

[Klein 1994] Bob Klein et Tim Stroup. *Grand Comics Database*, 1994. (Cité en page 49.)

[Klyne 2004] Graham Klyne et Jeremy J. Carroll. *Resource Description Framework (RDF) : Concepts and Abstract Syntax.* Rapport technique, 2004. (Cité en page 17.)

[Kuznetsov 2002] Sergei O. Kuznetsov et A. Obiedkov, Sergei. *Comparing performance of algorithms for generating concept lattices.* Journal of Experimental & Theoretical Artificial Intelligence, vol. 14, no. 2-3, pages 189–216, 2002. (Cité en page 137.)

[Lacassin 1971] Francis Lacassin. Pour un neuvième art, la bande dessinée. 1971. (Cité en page 1.)

[Lainé 2007] Jean-Marc Lainé et Sylvain Delzant. Le dessin des planches. Eyrolles, 2007. (Cité en page 36.)

[Lainé 2010] Jean-Marc Lainé et Sylvain Delzant. Le lettrage des bulles. Eyrolles, 2010. (Cité en pages 41, 43, 44 et 45.)

[Lamiroy 2014a] Bart Lamiroy et Jean-Marc Ogier. *Analysis and Interpretation of Graphical Documents.* In David Doermann et Karl Tombre, editeurs, Handbook of Document Image Processing and Recognition, pages 553–590. Springer London, 2014. (Cité en page 49.)

[Lamiroy 2014b] Bart Lamiroy et Shabai Zheng. *An Attempt to Use Ontologies for Document Image Analysis.* In RISE 2014 6eme Atelier Recherche d'Information SEmantique, 2014. (Cité en page 49.)

[Le Ber 2002] Florence Le Ber et Amedeo Napoli. *Design and comparison of lattices of topological relations based on Galois lattice theory.* In International Conference on Principles of Knowledge Representation and Reasoning, 2002. (Cité en page 60.)

[Leivant 1994] Daniel Leivant. *Higher Order Logic.* In In Handbook of Logic in Artificial Intelligence and Logic Programming, pages 229–321. Clarendon Press, 1994. (Cité en page 26.)

[Letz 1999] Reinhold Letz. *First-order Tableau Methods.* In Handbook of Tableau Methods, pages 125–196. 1999. (Cité en page 27.)

[Levesque 2007] Hector J. Levesque et Gerhard Lakemeyer. *Cognitive Robotics.* In Handbook of Knowledge Representation, pages 869–886. 2007. (Cité en page 34.)

[Li 2012a] Congcong Li, Adarsh Kowdle, Ashutosh Saxena et Tsuhan Chen. *Towards Holistic Scene Understanding : Feedback Enabled Cascaded Classification Models.* IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 7, pages 1394–1408, 2012. (Cité en page 49.)

[Li 2012b] Luyuan Li, Yongtao Wang, Zhi Tang et Liangcai Gao. *Automatic comic page segmentation based on polygon detection.* Multimedia Tools and Applications, vol. 69, no. 1, pages 171–197, 2012. (Cité en page 46.)

[Li 2013] Luyuan Li, Wang Zhi Tang, Xiaoqing Lu et Liangcai Gao. *Unsupervised speech text localization in comic images.* In International Conference on Document Analysis and Recognition (ICDAR), pages 1190–1194, 2013. (Cité en page 48.)

[Lindig 2002] Christian Lindig. *Fast Concept Analysis.* Working with Conceptual Structures-Contributions to ICCS, pages 235–248, 2002. (Cité en page 138.)

[Liu 2007] Ying Liu, Dengsheng Zhang, Guojun Lu et Wei-Ying Ma. *A survey of content-based image retrieval with high-level semantics.* Pattern Recognition, vol. 40, no. 1, pages 262–282, 2007. (Cité en page 56.)

[Maillot 2004] Nicolas Maillot, Monique Thonnat et Alain Boucher. *Towards ontology-based cognitive vision.* Machine Vision and Applications, vol. 16, no. 1, pages 33–40, 2004. (Cité en page 55.)

[Mao 2003] Song Mao, Azriel Rosenfeld et Tapas Kanungo. *Document Structure Analysis Algorithms : A Literature Survey.* Proceedings of SPIE - DRR, vol. 5010, pages 197–207, 2003. (Cité en page 49.)

[Marszalek 2007] Marcin Marszalek et Cordelia Schmid. *Semantic Hierarchies for Visual Object Recognition.* In CVPR, 2007. (Cité en page 54.)

[Massé 2006] Rodolphe Massé. Guide phénix du manga. Asuka édition, 2006. (Cité en page 4.)

[Mateu-Mestre 2010] Marcos Mateu-Mestre. Framed Ink, Drawing and composition for Visual Storytelling. Design Studio Press, 2010. (Cité en page 38.)

[McCloud 1993] Scott McCloud. Understanding Comics - The Invisible Art. 1993. (Cité en pages 2, 36 et 86.)

[McCloud 2006] Scott McCloud. Making Comics. Harper, 2006. (Cité en pages 36, 39, 41 et 46.)

[McGuinness 2004] Deborah L. McGuinness et Frank Van Harmelen. *OWL Web Ontology Language Overview.* Rapport technique, W3C, 2004. (Cité en pages 19 et 22.)

[McIntosh 2011] Jason McIntosh. *ComicsML*, 2011. (Cité en pages 49 et 50.)

[McLuhan 1968] Marshall McLuhan. Pour comprendre les médias. 1968. (Cité en page 36.)

[Mezaris 2003] Vasileios Mezaris, Ioannis Kompatsiaris et Michael G. Strintzis. *An Ontology Approach to Object-Based Image Retrieval.* ICIP, pages 511–514, 2003. (Cité en pages ix, 56, 57 et 58.)

[Miles 2005] Alistair Miles, Dan Brickley, Brian Matthews et Michael Wilson. *SKOS Core : Simple Knowledge Organisation for the Web.* In International Conference on Dublin Core and Metadata Applications, numéro 5, pages 3–10, 2005. (Cité en page 53.)

[Miller 1990] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross et Katherine J. Miller. *Introduction to WordNet : An On-line Lexical Database.* International Journal of Lexicography, vol. 3, no. 4, pages 235–244, 1990. (Cité en page 54.)

[Minsky 1974] Marvin Minsky. *A framework for representing knowledge.* Artificial intelligence, 1974. (Cité en page 23.)

[Miron 2007] Alina Dia Miron, Jérôme Gensel, Marlène Villanova-Oliver et Hervé Martin. *Relations spatiales qualitatives dans les ontologies géographiques avec ONTOAST.* SAGEO, 2007. (Cité en pages 58 et 59.)

[Morozumi 2009] Ayako Morozumi, Satomi Nomura, Mitsuharu Nagamori et Shigeo Sugimoto. *Metadata Framework for Manga : A Multi-paradign Metadata Description Framework for Digital Comics.* In International Conference on Dublin Core and Metadata Applications, pages 61–70, 2009. (Cité en page 50.)

[Morris 1964] Morris et Pierre Vankeer. *9e Art - Musée de la bande dessinée.* Spirou, 1964. (Cité en page 1.)

[Munro 1971] I. Munro. *Efficient determination of the transitive closure of a directed graph.* Information Processing Letters, pages 56–58, 1971. (Cité en page 138.)

[Norris 1978] E. Norris. *An algorithm for computing the maximal rectangles in a binary relation.* Revue roumaine de mathématiques appliquées, vol. 23, no. 2, 1978. (Cité en page 137.)

[Nourine 1999] Lhouari Nourine et Olivier Raynaud. *A fast algorithm for building lattices.* Information Processing Letters, vol. 71, no. 5-6, pages 199–204, 1999. (Cité en page 137.)

[ODLM 2014] ODLM. *Etat des lieux de l'offre de musique numérique.* Rapport technique, Hadopi, 2014. (Cité en page 5.)

[Open GIS Consortium Inc. 2011] Open GIS Consortium Inc. OpenGIS Implementation Standard for Geographic informations - Simple feature access - Part 1 : Common Architecture. 2011. (Cité en page 67.)

[O'Reilly 2007] Tim O'Reilly. *What is Web 2.0 : Design patterns and business models for the next generation of software.* Communications & Strategies, vol. 65, 2007. (Cité en page 14.)

[Paez 2008] Melissa Paez. *La bande dessinée japonaise. Comment le manga a-t-il séduit les Japonais et les francophones ?* Rapport technique, 2008. (Cité en page 4.)

[Pastierovic 2011] Robert Pastierovic. *Advanced Comic Book Format*, 2011. (Cité en page 50.)

[Petridis 2006] Kosmas Petridis, Dionysios Anastasopoulos, Carsten Saathoff, Norman Timmermann, Yiannis Kompatsiaris et Steffen Staab. *M-OntoMat-Annotizer : Image Annotation Linking Ontologies and Multimedia Low-Level Features.* In KES, pages 633–640, 2006. (Cité en page 58.)

[Pissavy-Yvernault 2013] Bertrand Pissavy-Yvernault et Christelle Pissavy-Yvernault. La véritable histoire de Spirou. Dupuis, 2013. (Cité en page 8.)

[Ponsard 2012] Christophe Ponsard, Ravi Ramdoyal et Daniel Dziamski. *An OCR-Enabled Digital Comic Books Viewer.* Computers Helping People with Special Needs, vol. 7382, pages 471–478, 2012. (Cité en page 134.)

[Prasad 2005] Mukul R. Prasad, Armin Biere et Aarti Gupta. *A survey of recent advances in SAT-based formal verification.* International Journal on Software Tools for Technology Transfer, vol. 7, no. 2, pages 156–173, 2005. (Cité en page 24.)

[Priss 2000] Uta Priss. *Lattice-based Information Retrieval.* Knowledge Organization, vol. 37, no. 3, pages 132–142, 2000. (Cité en page 136.)

[Prud'Hommeaux 2008] Eric Prud'Hommeaux et Andy Seaborne. *SPARQL : Query Langage for RDF.* Rapport technique, 2008. (Cité en page 19.)

[Quillian 1966] M. Ross Quillian. *Semantic Memory.* Semantic Information Processing, 1966. (Cité en page 23.)

[Ratier 2013] Gilles Ratier. *2013 : l'année de la décélération.* Rapport technique, Association des Critiques et journalistes de Bande Dessinée, 2013. (Cité en page 6.)

[Renouf 2007] Arnaud Renouf, Régis Clouard et Marinette Revenu. *How to formulate image processing applications ?* In International Conference on Computer Vision Systems, 2007. (Cité en page 61.)

[Richer 2008] Jean-Michel Richer. *Logique et Calcul Propositionnel*, 2008. (Cité en page 24.)

[Rigaud 2012] Christophe Rigaud, Norbert Tsopze, Jean-Christophe Burie et Jean-Marc Ogier. *Extraction robuste des cases et du texte de bandes dessinées.* In Colloque International Francophone sur l'Ecrit et le Document (CIFED), volume X, pages 349–360, 2012. (Cité en page 47.)

[Rigaud 2013a] Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier et Dimosthenis Karatzas. *Speech balloon contour classification in comics.* In International Workshop on Graphics Recognition (GREC), pages 23–25, 2013. (Cité en page 47.)

[Rigaud 2013b] Christophe Rigaud, Jean-Christophe Burie, Jean-Marc Ogier, Dimosthenis Karatzas et Joost Van De Weijer. *An Active Contour Model for Speech Balloon Detection in Comics.* In International Conference on Document Analysis and Recognition (ICDAR), pages 1240–1244. Ieee, 2013. (Cité en pages 47 et 120.)

[Rigaud 2013c] Christophe Rigaud, Dimosthenis Karatzas, Joost Van De Weijer, Jean-christophe Burie et Jean-marc Ogier. *Automatic Text Localisation in Scanned Comic Books.* In International Conference on Computer Vision Theory and Applications, pages 814–819, 2013. (Cité en pages 48 et 120.)

[Rigaud 2014a] Christophe Rigaud. *Segmentation and indexation of complex object in comic books.* PhD thesis, 2014. (Cité en pages 8, 46, 103, 108, 114, 120 et 149.)

[Rigaud 2014b] Christophe Rigaud et Clément Guérin. *Localisation contextuelle des personnages de bandes dessinées.* In Colloque International Francophone sur l'Ecrit et le Document (CIFED), 2014. (Cité en page 120.)

[Rigaud 2014c] Christophe Rigaud, Dimosthenis Karatzas, Jean-Christophe Burie et Jean-Marc Ogier. *Color descriptor for content-based drawing retrieval.* In International Workshop on Document Analysis Systems (DAS), pages 267–271, 2014. (Cité en pages 48 et 114.)

[Rissen 2012] Paul Rissen. *A Comics Ontology*, 2012. (Cité en pages x, 51, 150 et 177.)

[Rohloff 2007] Kurt Rohloff, Mike Dean, Ian Emmons, Dorene Ryder et John Sumner. *An Evaluation of Triple-Store Technologies for Large Data Stores.* Lecture Notes in Computer Science - On the Move to Meaningful Internet Systems, vol. 4806, pages 1105–1114, 2007. (Cité en page 19.)

[Rozière 2011] Paul Rozière. *Logique mathématique : une introduction au calcul des prédicats du premier ordre*, 2011. (Cité en page 26.)

[Schmidt-Schauss 1991] Manfred Schmidt-Schauss et Gert Smolka. *Attributive concept descriptions with complements.* Artificial intelligence, vol. 48, no. 1, pages 1–26, 1991. (Cité en pages 28, 29 et 32.)

[Schneider 2003] Luc Schneider. *Designing Foundational Ontologies.* Conceptual Modeling - ER 2003, vol. LNCS 2813, pages 91–104, 2003. (Cité en page 20.)

[Schreiber 2001] A. Th. Guus (Guus) (Guus) Schreiber, Barbara Dubbeldam, Jan Wielemaker et Bob Wielinga. *Ontology-Based Photo Annotation.* IEEE Intelligent Systems, vol. 16, no. 3, pages 66–74, 2001. (Cité en page 54.)

[Searle 1980] John R. Searle. *Minds, brains, and programs.* Behavioral and Brain Sciences, vol. 3, no. 3, 1980. (Cité en page 17.)

[Shadbolt 2006] Nigel Shadbolt, Wendy Hall et Tim Berners-Lee. *The Semantic Web Revisited.* IEEE Intelligent Systems, vol. 21, no. 3, pages 96–101, 2006. (Cité en page 19.)

[Sikora 2001] Thomas Sikora. *The MPEG-7 Visual Standard for Content Description — An Overview.* IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, pages 696–702, 2001. (Cité en page 58.)

[Smeulders 2000] Arnold W.M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta et Ramesh Jain. *Content-based image retrieval at the end of the early years.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pages 1349–1380, 2000. (Cité en pages 55 et 56.)

[Smullyan 1968] Raymond M. Smullyan. First-Order Logic. Springer, 1968. (Cité en page 27.)

[Staab 2009] Steffen Staab et Rudi Studer. Handbook on ontologies. Springer, 2009. (Cité en page 33.)

[Stevens 2009] Robert Stevens et Phillip Lord. *Application of Ontologies in Bioinformatics.* In Handbook on Ontologies, pages 735–756. 2009. (Cité en page 33.)

[Stocker 2009] Markus Stocker et Evren Sirin. *PelletSpatial : A Hybrid RCC-8 and RDF / OWL Reasoning and Query Engine.* In Web Ontology Language : Experience and Directions (OWLED), volume 2009. Springer Verlag, 2009. (Cité en page 60.)

[Studer 1998] Rudi Studer, V. Richard Benjamins et Dieter Fensel. *Knowledge Engineering : Principles and Methods.* Data & Knowledge Engineering, vol. 25, no. 1-2, pages 161–198, 1998. (Cité en page 20.)

[Stumme 2002] Gerd Stumme, Rafik Taouil, Yves Bastide, Nicolas Pasquier et Lotfi Lakhal. *Computing iceberg concept lattices with TITANIC.* Data & Knowledge Engineering, vol. 42, no. 2, pages 189–222, 2002. (Cité en page 140.)

[Su 2011] Chung-Yuan Su, Ray-I Chang et Jen-Chang Liu. *Recognizing text elements for svg comic compression and its novel applications.* In International Conference on Document Analysis and Recognition (ICDAR), pages 1329–1333, 2011. (Cité en page 48.)

[Sun 2013] Weihan Sun et Koichi Kise. *Detection of exact and similar partial copies for copyright protection of manga.* International Journal on Document Analysis and Recognition (IJDAR), vol. 16, no. 4, pages 331–349, 2013. (Cité en page 48.)

[Sundaresan 2012] M. Sundaresan et S. Ranjini. *Text extraction from digital english comic image using two blobs extraction method.* In International Conference on Pattern Recognition, Informatics and Medical Engineering, pages 449–452, 2012. (Cité en page 48.)

[Takayama 2012] Kohei Takayama, Henry Johan et Tomoyuki Nishita. *Face detection and face recognition of cartoon characters using feature extraction.* In Image, Electronics and Visual Computing Workshop, 2012. (Cité en page 48.)

[Text Encoding Initiative Consortium 2014] Text Encoding Initiative Consortium. *Text Encoding Initiative*, 2014. (Cité en page 50.)

[The Gene Ontology Consortium 2000] The Gene Ontology Consortium. *Gene Ontology : tool for the unification of biology.* Nature Genetics, vol. 25, no. may, pages 25–29, 2000. (Cité en page 33.)

[Tousch 2012] Anne-Marie Tousch, Stéphane Herbin et Jean-Yves Audibert. *Semantic hierarchies for image annotation : a survey.* Pattern Recognition, vol. 45, no. 1, pages 333–345, 2012. (Cité en pages 34 et 55.)

[Uschold 1995] Mike Uschold et Martin King. *Towards a Methodology for Building Ontologies.* In Workshop on Basic Ontological Issues in Knowledge Sharing, numéro July, 1995. (Cité en page 20.)

[Uschold 1996] Mike Uschold et Michael Gruninger. *Ontologies : Principles, methods and applications.* The Knowledge Engineering Review, vol. 11, no. 02, pages 93–136, 1996. (Cité en page 19.)

[Van der Hoek 2007] Wiebe Van der Hoek et Michael Wooldridge. *Multi-Agent Systems.* In Handbook of Knowledge Representation, pages 887–928. 2007. (Cité en page 34.)

[Van Harmelen 2007] Frank Van Harmelen, Vladimir Lifschitz et Bruce Porter. Handbook of Knowledge Representation. Elsevier, 2007. (Cité en page 33.)

[W3C 2012a] W3C. *OWL 2 Web Ontology Language Direct Semantics.* Rapport technique, 2012. (Cité en page 30.)

[W3C 2012b] W3C. *OWL 2 Web Ontology Language Document Overview.* Rapport technique, 2012. (Cité en page 22.)

[W3C 2012c] W3C. *OWL 2 Web Ontology Language Profiles.* Rapport technique, 2012. (Cité en page 22.)

[W3C 2012d] W3C. *Periodical Comics*, 2012. (Cité en page 49.)

[Walsh 2012] John A Walsh. *Comic Book Markup Language : An Introduction and Rationale.* Digital Humanities Quarterly (DHQ), vol. 6, no. 1, pages 1–50, 2012. (Cité en page 50.)

[Wille 1982] Rudolf Wille. *Restructuring Lattice Theory : An Approach Based on Hierarchies of Concepts.* In Ordered Sets, pages 445–470. Dordrecht-Boston : Reidel, i. rival édition, 1982. (Cité en page 137.)

[Yamada 2004] Masashi Yamada et Rahmat Budiarto. *Comic image decomposition for reading comics on cellular phones.* IEICE, no. xx, pages 1–8, 2004. (Cité en page 48.)

[Yao 2007] Benjamin Yao, Xiong Yang et Song-Chun Zhu. *Introduction to a Large-Scale General Purpose Ground Truth Database : Methodology, Annotation Tool and Benchmarks.* Energy Minimization Methods in Computer Vision and Pattern Recognition, vol. 4679, pages 169–183, 2007. (Cité en page 88.)

[Yao 2012] Jian Yao, Sanja Fidler et Raquel Urtasun. *Describing the Scene as a Whole : Joint Object Detection, Scene Classification and Semantic Segmentation.* In Computer Vision and Pattern Recognition (CVPR), pages 702–709, 2012. (Cité en page 49.)

[Zhang 2002] Lintao Zhang et Sharad Malik. *The Quest for Efficient Boolean Satisfiability Solvers.* In A. Voronkov, editeur, International Conference on Automated Deduction, pages 295–313. Springer, 2002. (Cité en page 24.)

[Zhang 2009] Dengsheng Zhang, Monirul Islam, Guojun Lu et Jin Hou. *Semantic Image Retrieval Using Region Based Inverted File.* In Digital Image Computing : Techniques and Applications, pages 242–249, 2009. (Cité en page 58.)

[Zinger 2005] S. Zinger, C. Millet, B. Mathieu, G. Grefenstette, P. Hède et P.-A. Moëllic. *Extracting an Ontology of Portrayable Objects from WordNet.* In

MUSCLE/ImageCLEF workshop on Image and Video retrieval evaluation, pages 17–23, 2005. (Cité en pages 54, 78 et 152.)

[Zouari 2009] Khaled Zouari. *Le site Webcomics, un exemple d'interactivité.* Hermès, La Revue, no. 54, pages 127–132, 2009. (Cité en pages 5 et 7.)

Annexe A

Annexes

A.1 Protocole de construction de la vérité terrain

A.1.1 Pré-traitement

Il a été décidé de ne plus effectuer de pré-traitement avant la construction de la VT. D'une part le logiciel de saisie a grandement gagné en ergonomie ce qui rend le travail beaucoup moins fastidieux qu'il a pu l'être précédemment. D'autre part, partir d'une page vierge nous assure qu'aucun biais en faveur de l'algorithme de pré-traitement ne sera intégré à la VT.

A.1.2 Cases

Un polygone doit être construit autour de chaque case de la page. Le polygone doit englober tous les pixels de la case. Le cadrage doit se faire « au plus proche des pixels mais sans les toucher ». Cases délimitées par un cadre : Tout objet (bulle, bout de personnage, etc.) dépassant hors du cadre doit être tronqué. Le polygone doit suivre la géométrie du cadre, si la case est triangulaire, le polygone doit être un triangle, si la case est un cercle, le polygone doit être une approximation de ce cercle. Deux cases se chevauchant doivent être détourées comme deux cases distinctes Si une case est incluse dans une autre, les deux cases doivent être détourées de manière distincte Cases non-délimitées par un cadre : Le polygone doit représenter le cadre que pourrait avoir la case. Il doit cependant continuer à respecter la règle du « au plus proche mais sans toucher ».

A.1.3 Zones de texte

Le polygone doit être construit autour de chaque ligne de la zone de texte, toujours « au plus proche mais sans toucher les pixels». Dans la mesure du possible, aucun objet autre que le texte ne doit apparaître à l'intérieur du polygone. On parle bien ici du texte et non des bulles. Il faut donc détourer le texte présent dans les bulles mais également le texte inscrit directement dans l'image ainsi que les onomatopées.

A.2 Rappel sur l'utilisation de Toonshop

A.2.1 Créer une vérité terrain sur une page

Les polygones sont créés sur les SVG du dossier « GroundTruth ». Pour créer une VT sur une page, il faut faire un clic droit sur le fichier SVG correspondant puis sélectionner « Modifier cette page ». Vous entrez alors en mode édition où vous pouvez ajouter des polygones ou supprimer/modifier des polygones déjà existants. Vous pouvez également ajouter ou modifier les métadonnées associées à la page et aux polygones. Afin de pouvoir ajouter/modifier/supprimer des polygones, il faut que seul le calque correspondant au type de polygone souhaité soit affiché. L'affichage des calques se règle grâce aux boutons « Frame », « Text » et « Object » en haut de l'écran.

A.2.2 Créer un polygone

Pour créer un nouveau polygone, deux solutions : Clic droit : « Ajouter un polygone » pour créer un polygone rapidement. Un nouveau polygone sera alors ajouté au calque. Son point supérieur gauche aura pour coordonnées l'emplacement de la souris au moment du clic. Clic droit : « Créer un polygone "Point à Point" ». Le premier point sera placé à l'endroit où le clic droit aura été effectué et sera marqué par un cercle jaune. Il suffit ensuite de cliquer, avec le bouton gauche de la souris, sur l'image aux endroits où vous souhaitez que les prochains points soient placés. Pour finaliser le polygone, appuyez sur la touche « Entrée » ou sur le bouton droit de la souris, ne cliquez pas sur le premier point !

A.2.3 Modifier un polygone

Avant de modifier un polygone il faut d'abord le sélectionner en utilisant le clic gauche. Vous pouvez ensuite sélectionner un de ses points également avec le clic gauche. A noter que lorsque vous sélectionnez un polygone, son premier point est automatiquement sélectionné. Pour déselectionner un point ou un polygone, il suffit de cliquer une nouvelle fois dessus avec le bouton gauche de la souris. Lorsqu'un point est sélectionné, il est possible d'effectuer plusieurs actions : Déplacer un point Clic droit : « Déplacer le point ici » déplacera le point sélectionné à l'endroit voulu. Maintenir la touche « Alt » + une flèche de direction déplacera le point pixel à pixel dans la direction souhaitée. Notez qu'en maintenant également la touche « Shift » enfoncée le point se déplacera avec un pas de 10 pixels. Supprimer un point Clic droit : « Supprimer le point sélectionné » supprimera le point. Appuyer sur la touche « Suppr » supprimera le point. Ajouter un point Clic droit : « Ajouter un nouveau point ici » ajoutera un nouveau point à l'endroit cliqué, entre le point sélectionné et celui qui le suivait dans le polygone. Notez qu'il est possible de changer de point sélectionné en appuyant sur la touche « Tabulation ».

Lorsque toutes vos modifications sont effectuées, n'oubliez pas de cliquer sur le bouton « Enregistrer les modifications » sans quoi elles ne seront pas conservées.

A.2.4 Supprimer un polygone

Pour supprimer un polygone il faut d'abord le sélectionner avec le clic gauche, ensuite deux solutions : Clic droit : « Supprimer le polygone sélectionné » Appuyer sur la touche « Suppr »

Attention, par défaut le premier point est sélectionné, veillez à bien le déselectionner avant de continuer.

A.2.5 Utilisation du zoom et navigation

Il est possible de zoomer sur certains points de l'image lors de la création de la vérité terrain, pour cela plusieurs solutions : Maintenir la touche « Shift » et le clic droit enfoncés puis déplacer la souris vers le bas de l'écran augmente le facteur de zoom sur la zone visée. Maintenir la touche « Shift » et le clic droit enfoncés puis déplacer la souris vers le haut de l'écran diminue le facteur de zoom sur la zone visée. Maintenir la touche « Ctrl » et le clic gauche enfoncés puis définir un cadre avec la souris permet de zoomer sur ce cadre. Maintenir les touches « Ctrl » et « Shift » enfoncés puis effectuer un clic droit permet de revenir à un affichage pleine page.

Il est également possible de déplacer la zone d'affichage : Maintenir la touche « Shift » et le clic gauche enfoncés puis déplacer le curseur de la souris pour déplacer l'image dans le canvas.

A.2.6 Ajouter ou modifier des métadonnées

Page à page Il est possible d'ajouter des métadonnées sur les pages, les cases et les zones de texte. Lorsque vous éditez une page, le volet sur la droite de l'écran affiche les métadonnées disponibles pour le type de polygone sélectionné. Si aucun polygone n'est sélectionné, les métadonnées affichées sont celles de la page. Il est possible d'éditer ces métadonnées en inscrivant leur valeur directement dans les champs correspondants. Plusieurs pages à la fois Certaines métadonnées sont redondantes entre les pages (nom de l'album, de l'auteur, etc.). Il est possible de définir les mêmes métadonnées pour un ensemble de page en sélectionnant ces pages dans le volet de gauche. Il faut ensuite faire un clic droit puis choisir « Modifier les métadonnées ». Il suffit alors de compléter les champs puis de valider. Attention à bien finaliser l'écriture d'un champ en appuyant sur « Entrée » avant de fermer la fenêtre d'édition.

A.3 Formalismes de bande dessinée

A.3.1 ComicsML

Ci-dessous la DTD du formalisme XML ComicsML, présenté en 3.4.

```
<!-- Comics Markup Language -->
```

```
<!-- VERSION 0.2 -->

<!--
$Id: comics_ml.dtd,v 1.1.1.1 2002/03/12 03:01:45 jmac Exp $
-->

<!-- TODO
What to do about noises characters make? Grumbling, stuttering, spoken sound
    effects, etc.
Investigate Dublin Core namespace to use for all the authorship information.
-->

<!-- Central elements of this DTD:
comic - The concept of a comic, rather than the actual, readable
comics themselves. This holds all the meta-information about a
particular group of comics, and holds zero or more strip elements.
strip - One instance of the actual comic. Could be each episode in a
serial adventure, each strip in a newspaper-style gag strip, and so
on. A strip contains one or more panels.
panel - One image in a strip element. This is where the actual content
lives.
-->

<!-- Meta information about the comic in general -->
<!ELEMENT comic (title, person+, last-built, icon?, description?, url?, strip*)>

<!ATTLIST comic version CDATA #REQUIRED
               xml:lang CDATA #IMPLIED>

<!ELEMENT title (#PCDATA)>

<!-- Stuff about a person involved in making the comic -->
<!ELEMENT person (firstname?, surname, email?, url?, job?)>
<!ELEMENT firstname (#PCDATA)>
<!ELEMENT surname (#PCDATA)>
<!ELEMENT email (#PCDATA)>
<!-- person's homepage -->
<!ELEMENT url (#PCDATA)>
<!-- person's relationship to comic (writer, artist, caterer, etc.) -->
<!ELEMENT job (#PCDATA)>

<!-- timestamp of most recent strip -->
<!ELEMENT last-built (#PCDATA)>
<!-- comic's distributable logo (such as an ad banner) -->
<!ELEMENT icon (#PCDATA)>
<!-- description about this comic -->
<!ELEMENT description (#PCDATA)>

<!-- An instance of a comic -->
<!-- The optional 'person' element can be used to list people who
worked on this particular strip, but aren't in the regular credits -->
<!-- The 'url' element can point to where the strip exists in full -->
<!ELEMENT strip (date, title?, person*, url?, teaser?, panels?)>
<!-- Each strip must have a unique serial number -->
```

```
<!ATTLIST strip id ID #REQUIRED
              xml:lang CDATA #IMPLIED>

<!ELEMENT date (#PCDATA)>

<!ELEMENT panels (panel+)>
<!-- An image within a strip -->
<!-- The 'url' element points to where this panel is available -->
<!ELEMENT panel (url?, caption?, panel-desc*)>
<!ATTLIST panel xml:lang CDATA #IMPLIED>

<!-- Caption for a panel -->
<!ELEMENT caption (#PCDATA)>

<!-- The teaser briefly describes what happens in this strip, and may
be useful in synidcation -->
<!ELEMENT teaser (description?, panels?)>

<!-- Panel description elements -->

<!-- Some inlines for simple text markup -->
<!ENTITY \% inlines 'emphasis | strong | soft'>
<!--
'emphasis' - Ordinary spoken emphasis, which many comics make explicit
by emboldening the appropriate words.
'strong' - As above, but extra-strong emphasis, outside of normal
conversational tones.
'soft' - Speech that is distant or otherwise softer than normal
conversation, usually depicted as smaller lettering.
-->
<!ELEMENT emphasis (#PCDATA)>
<!ELEMENT strong (#PCDATA)>
<!ELEMENT soft (#PCDATA)>

<!-- More inlines for 'vocal effects', things characters can 'say' that aren't
    actually speech. -->
<!ENTITY \% vocal-effects 'interro | bang | interrobang | silence'>
<!--
'interro' - '?'
'bang' - '!'
'interrobang' - '?!' or something similar
'silence' - An empty balloon, or '...' by itself; the character is
emoting a palpable pause.
Note that these may be combined with other markup to adjust their
"volume". A character who is really surprised and says "!!!" might say
"<strong><bang/></strong> in ComicsML.
-->
<!ELEMENT interro EMPTY>
<!ELEMENT bang EMPTY>
<!ELEMENT interrobang EMPTY>
<!ELEMENT silence EMPTY>

<!-- Text -->
<!ELEMENT panel-desc (narration | thought | speech | action)*>
```

```
<!ATTLIST panel-desc xml:lang CDATA #IMPLIED>
<!--
narration - Narrative text not spoken by any character.
thought - Unspoken character thoughts, usually shown in a puffy
thought bubble.
speech - Character speech, as depicted in word balloons, which can
come in a variety of shapes; see the 'tone' attribute.
action - A description of what is visually happening in this panel.
-->
<!ELEMENT narration (#PCDATA | \%inlines; | \%vocal-effects;)*>
<!ELEMENT action (#PCDATA | \%inlines;)*>

<!-- Multiple characters can share a speech or thought bubble. -->
<!ELEMENT thought (character+, text)>
<!ELEMENT speech (character+, text)>

<!-- Speech bubbles come in many tones, which comics usually represent
through different bubble shapes. (The descriptions below are based on
Western comics tradition; comics in other cultures may use different
shapes, but the varieties of content, which is what the attribute
actually expresses, remains the same, since they all represent
universal human speaking styles. Aint XML cool?)

normal - Speech that would go in an ordinary, round word balloon. This
is the default attribute value.

shout - Louder-than-normal speech, as represented by a spiky-bordered
balloon.

whisper - Whispered speech, often drawn in a balloon with a
broken-line border.

cold - A knife-edged utterance from a likely angry character, as found
within a balloon with icicle outlines hanging from it.

wobbly - Wavering, unsteady speech from an upset character (as during
crying, vomiting, and other speech-hindering acts), usually shown
inside a ballon with liquid, wavy-lined borders.
-->

<!-- The 'target' attribute lists whom this speech is directed at, if
it isn't obvious. Rarely used, but useful if a character, for example,
talks to another character in the room while holding a simultaneous
conversation with someone on a telephone, and all in the same panel. -->
<!ATTLIST speech tone (shout|whisper|cold|wobbly|normal) #IMPLIED
               target CDATA #IMPLIED>
<!-- The text of the character's speech -->
<!ELEMENT text (#PCDATA | \%inlines; | %vocal-effects;)*>
<!ATTLIST text xml:lang CDATA #IMPLIED>
<!-- The name of the speaking chracater -->
<!ELEMENT character (#PCDATA)>
```

ENCADRE A.1 – DTD de ComicsML

A.3.2 A Comics Ontology

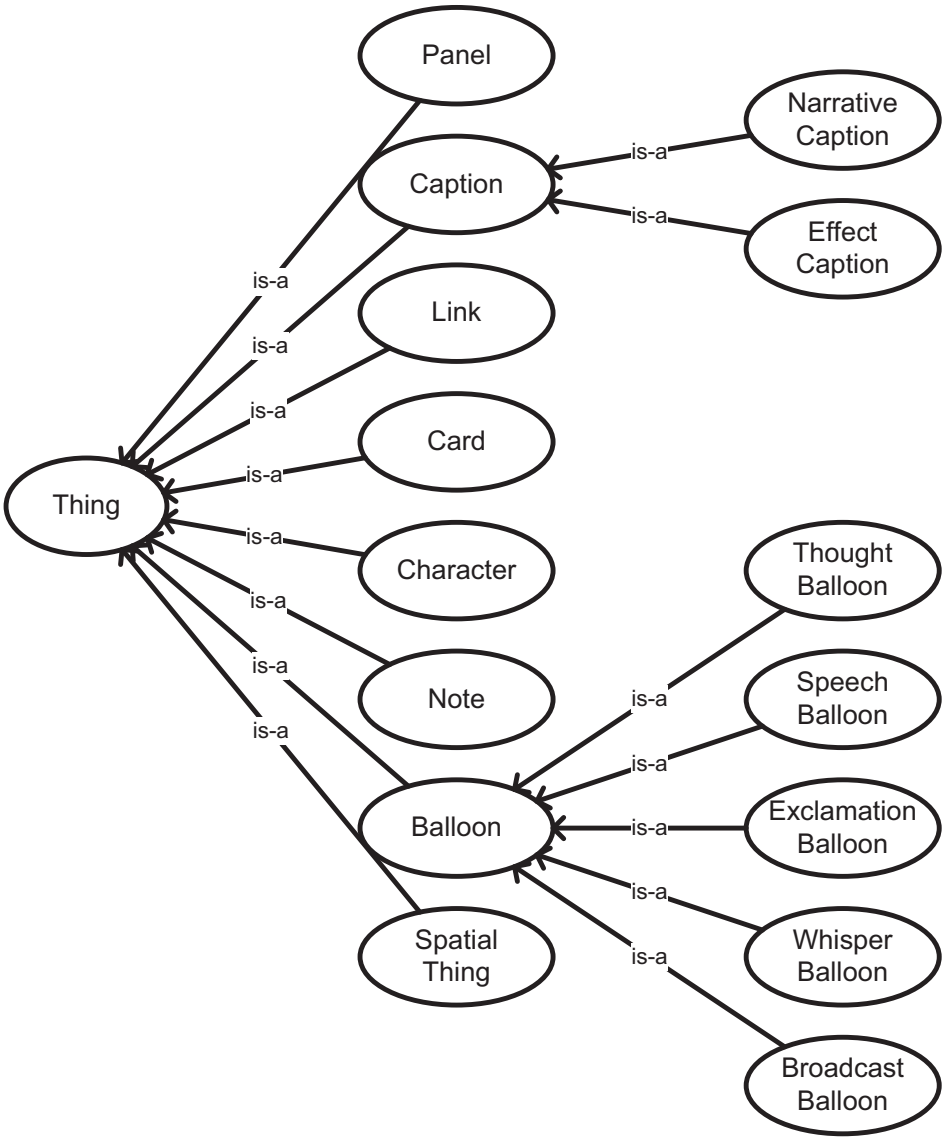


FIGURE A.1 – Hiérarchie de concepts de l'ontologie proposée par [Rissen 2012]

Propriété	Description
balloon	A panel can have one or many balloons inside it.
balloonOrder	This property assigns an order to a Balloon within a Panel.
broadcast-content	This property links a broadcast balloon to its content.
caption	This property allows a Card or Panel to have a Caption attached.
character	A panel can feature one or more characters.
directed-at	This property links a balloon to a character or object being addressed in the content.
effect-content	This property links an effect caption to its content.
exclamation-content	This property links an exclamation balloon to its content.
has-link	This property links a note to a URL of a primary or secondary source which provides background evidence.
has-note	This property links a balloon or panel to a explanatory or side note.
hasPanel	A Card can have one or many Panels inside it.
in-reply-to	This property links a balloon to another balloon which has content being referred to in the content.
linkedBalloon	This property links one Balloon to another, e.g. to suggest they are part of the same speech.
location	This property gives the events portrayed in a Card or Panel a location in the story world.
narrative-content	This property links a narrative caption to its content.
note-content	This property links a note container to the contents of the note.
onPanel	A Caption can be placed upon one or more Panels.
panelOrder	This property assigns an order to a Panel on a Card.
source	This property links a balloon or a caption to the character or object from which it emerges.
speech-content	This property links a speech balloon to its content.
thought-content	This property links a thought balloon to its content.
whisper-content	This property links a whisper balloon to its content.

TABLEAU A.1 – Description des propriétés de l'ontologie proposée par Paul Rissen

Proposition d'un cadre pour l'analyse automatique, l'interprétation et la recherche interactive d'images de bande dessinée.

Résumé : Le paysage numérique de la culture française et mondiale subit de grands bouleversements depuis une quinzaine d'années avec des mutations historiques des médias, de leur format traditionnel au format numérique, tirant avantageusement parti des nouveaux moyens de communication et des dispositifs mobiles aujourd'hui popularisés. Aux côtés de formes culturelles ayant achevé, ou étant en passe d'achever, leur transition vers le numérique, la bande dessinée tâtonne encore pour trouver sa place dans l'espace du tout dématérialisé. En parallèle de l'émergence de jeunes auteurs créant spécifiquement pour ces nouveaux supports de lecture que sont ordinateurs, tablettes et smartphones, plusieurs acteurs du monde socio-économique s'intéressent à la valorisation du patrimoine existant. Les efforts se concentrent autant sur une démarche d'adaptation des œuvres aux nouveaux paradigmes de lecture que sur celle d'une indexation de leur contenu facilitant la recherche d'informations dans des bases d'albums numérisés ou dans des collections d'œuvres rares. La problématique est double, il s'agit premièrement d'être en mesure d'identifier la structure d'une planche de bande dessinée en se basant sur des extractions de primitives, issues d'une analyse d'image, validées et corrigées grâce à l'action conjointe de deux ontologies, la première manipulant les extractions d'images bas-niveau, la deuxième modélisant les règles de composition classiques de la bande dessinée franco-belge. Dans un second temps l'accent est mis sur l'enrichissement sémantique des éléments identifiés comme composants individuels d'une planche en s'appuyant sur les relations spatiales qu'ils entretiennent les uns avec les autres ainsi que sur leurs caractéristiques physiques intrinsèques. Ces annotations peuvent porter sur des éléments seuls (place d'une case dans la séquence de lecture) ou sur des liens entre éléments (texte prononcé par un personnage).

Mots clés : bande dessinée, images, données complexes, représentation des connaissances, ontologies, raisonnement spatial, fossé sémantique, recherche d'informations, interactivité

A framework for the automated analysis, interpretation and interactive retrieval of comic books' images.

Abstract: Since the beginning of the twenty-first century, the cultural industry, both in France and worldwide, has been through a massive and historical mutation. They have had to adapt to the emerging digital technology represented by the Internet and the new handheld devices such as smartphones and tablets. Although some industries successfully transfered a piece of their activity to the digital market and are about to find a sound business model, the comic books industry keeps looking for the right solution and has not yet produce anything as convincing as the music or movie offers. While many new young authors and writers use their creativity to produce specifically digital designed pieces of art, some other minds are focused on the preservation and the development of the already existing heritage. So far, efforts have been concentrated on the transfer from printed to digital support, with a special attention given to their specific features and how they can be used to create new reading conventions. There has also been some concerns about the content indexing, which is a hard task regarding the large amount of data created since the very beginning of the comics history. From a scientific point of view, there are several issues related to these goals. First, it implies to be able to identify the underlying structure of a comic books page. This comes through the extraction of the page's components, their validation and their correction based on the representation and reasoning capacities of two ontologies. The first one focus on the representation of the image analysis concepts and the second one represents the comic books domain knowledge. Secondly, a special attention is given to the semantic enhancement of the extracted elements, based on their spatial relations to each others and on their own characteristics. These annotations can be related to elements only (e.g. the position of a panel in the reading sequence), or to the bound between several elements (e.g. the text pronounced by a character).

Keywords: comic books, images, complex data, knowledge representation, ontologies, spatial reasoning, semantic gap, information retrieval, interactivity

Laboratoire Informatique, Image, Interaction
Pôle Sciences & Technologie
Avenue Michel Crépeau

17042 LA ROCHELLE CEDEX 1