

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité

INFORMATIQUE MÉDICALE

Présentée par

M. Jean-Baptiste LAMY

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PARIS 6

Conception et évaluation de méthodes de
visualisation des connaissances médicales :

Mise au point d'un langage graphique et
application aux connaissances sur le médicament

Directeur de thèse : Alain VENOT
Co-encadrante : Catherine DUCLOS

Soutenue publiquement le : 19 / 12 / 2006

Devant le jury composé de :

Pr. Alain VENOT, directeur de thèse

Pr. Stéfan DARMONI et Pr. Marius FIESCHI, rapporteurs

M. Pierre ZWEIGENBAUM et Pr. Alain-Jacques VALLERON, examinateurs

“There is no field where humanisation of knowledge through the eye would not be possible.”
– Otto Neurath, *Autobiography*

Remerciements

Je remercie Alain Venot pour avoir dirigé ma thèse et pour son accueil chaleureux au Laboratoire d’Informatique Médicale et BIOinformatique (LIM&BIO, UFR SMBH, Université Paris XIII).

Je remercie Catherine Duclos pour ses idées et sa rigueur, ainsi que pour avoir co-encadré ce travail de thèse.

Je remercie Stéphane Darmoni et Marius Feschi, qui ont accepté d’être rapporteurs et de consacrer du temps à l’évaluation de ce travail.

Je remercie Pierre Zweigenbaum et Alain-Jacques Valleron d’avoir accepté de participer au jury de ma thèse, en tant qu’examinateur.

Je remercie Vincent Rialle pour ses conseils en matière de science cognitive et pour son amitié.

Je remercie Jean-Daniel Zucker pour les idées que nous avons pu échanger sur la grammaire du langage VCM.

Je remercie Patrick Ouvrard et les médecins évaluateurs de la SFTG pour le temps qu’ils ont consacré à VCM, pour leur propositions et pour leur enthousiasme.

Je remercie Avner Bar-Hen pour son aide lors de l’analyse statistique des données recueillies au cours de l’évaluation de VCM.

Je remercie Vincent Danel, Jean Calop et Bernard Champon pour avoir consacré une partie de leur temps à suivre ce travail.

Je remercie tous les membres du LIM&BIO pour leur accueil chaleureux.

Je remercie la CNAMTS pour avoir subventionné l’évaluation du langage VCM.

Je remercie toute ma famille et mes amis pour leur soutien.

Table des matières

1	Introduction	15
1.1	Problématique	15
1.2	Objectifs	16
1.2.1	VCM et le langage naturel	17
1.3	Plan de la thèse	18
2	État de l’art des modèles de la vision, des langages graphiques et de la visualisation d’information	19
2.1	La sémiotique et les modèles de la vision	19
2.1.1	Sémiotique	20
2.1.2	Psychologie de la forme (théorie de la <i>Gestalt</i>)	20
2.1.3	Sémiologie graphique	21
2.1.4	Modèle neurophysiologique	23
2.1.5	Perception pré-attentive	23
2.1.6	Conclusion sur les modèles de la vision	24
2.2	Les langages graphiques	25
2.2.1	Écritures graphiques	25
2.2.1.1	Hiéroglyphes égyptiens	25
2.2.1.2	Sinogrammes	26
2.2.1.3	Langages graphiques “universels”	26
2.2.2	Pictogrammes standardisés ou recommandés	27
2.2.2.1	Le jeu de pictogrammes général de l’ISO	27
2.2.3	Pictogrammes et langages graphiques dans la vie de tous les jours	27
2.2.3.1	Signalisation routière	27
2.2.3.2	Étiquetage des produits chimiques	29
2.2.3.3	Icônes de Logiciels et d’Environnements Informatiques Standardisés (ILEIS)	29
2.2.4	Pictogrammes et langages graphiques dans le domaine médical	31
2.2.4.1	Le Symbol sourcebook	31
2.2.4.2	Pictogrammes de l’AFSSAPS	32
2.2.4.3	Icônes du projet OPADE	32
2.2.4.4	Stabilis 2	32
2.2.4.5	Pictogrammes de l’USP	33
2.2.4.6	UVAL-MED	33
2.2.5	Conclusion sur les langages graphiques	34
2.3	La visualisation d’information	35
2.3.1	<i>Fisheye</i>	35
2.3.2	Méthodes de visualisation de textes et de listes	36
2.3.2.1	Repliement (<i>foldng</i>)	36
2.3.2.2	<i>Fisheye</i> déformant	36
2.3.2.3	<i>Greeking</i>	36

2.3.3	Méthodes de visualisation de données multidimensionnelles et matrices objet-attribut	38
2.3.3.1	Graphiques	38
2.3.3.2	Coordonnées parallèles	39
2.3.3.3	Glyphes	39
2.3.3.4	Lentille de tableau	41
2.3.3.5	Réduction réciproque de liste	42
2.3.3.6	Requêtes	43
2.3.4	Méthodes de visualisation d'arbres	43
2.3.4.1	Diagrammes	43
2.3.4.2	Arbres hyperboliques	45
2.3.4.3	Arbres coniques (<i>cone tree</i>)	45
2.3.4.4	Cartes d'arbres (<i>tree map</i>)	45
2.3.4.5	Arbres parallèles (<i>overlapping hierarchies</i>)	46
2.3.5	Méthodes de visualisation de réseaux et graphes	47
2.3.5.1	Diagrammes	47
2.3.5.2	Multi-arbres	47
2.3.5.3	<i>Fisheye</i> déformant	47
2.3.6	Méthodes de visualisation par indices de similarité	47
2.3.6.1	Terrains 3D	47
2.3.6.2	Système multi-agents	48
2.3.6.3	Points d'intérêt	49
2.3.7	Métaphores	49
2.3.7.1	Arbre botanique	50
2.3.7.2	Villes	50
2.3.8	Conclusion sur la visualisation d'information	50
2.4	Conclusion : quelle approche graphique pour les connaissances sur le médicament ?	53
3	État de l'art des connaissances sur le médicament	54
3.1	Définitions de la connaissance	54
3.2	Les monographies des médicaments	55
3.2.1	Les monographies sous forme textuelle	55
3.2.2	Modèles structurés des monographies	55
3.2.2.1	Modèles XML	56
3.2.2.2	Indications	56
3.2.2.3	Contre-indications	57
3.2.2.4	Précautions d'emploi et mises en garde	57
3.2.2.5	Interactions médicamenteuses	57
3.2.2.6	Effets indésirables	57
3.2.2.7	Pharmacocinétique	58
3.3	Bases de connaissances médicamenteuses	58
3.4	Guide de bonnes pratiques	60
3.4.1	Les guides de bonnes pratiques sous forme textuelle	60
3.4.2	Modèles structurés des GBP	61
3.5	Classifications médicales	61
3.5.1	CIM10	62
3.5.2	SNOMED RT	62
3.5.3	MedDRA	62
3.5.4	ATC	62
3.5.5	CCAM	62
3.5.6	UMLS	63
3.6	Besoins en connaissance sur le médicament	63

3.6.1	Besoins des professionnels de santé	63
3.6.2	Besoins des patients	64
3.7	Conclusion	64
4	Modélisation des connaissances sur le médicament	66
4.1	Introduction et objectifs	66
4.2	Matériels et méthodes	66
4.2.1	Connaissances d'expert du domaine	67
4.2.2	Modèles existants	67
4.2.3	Analyse d'un corpus de RCP	68
4.2.3.1	Matériels	68
4.2.3.2	Analyse des attributs, leurs fréquences et de leurs valeurs	68
4.2.3.3	Analyse de la structure des phrases	70
4.3	Résultats	71
4.3.1	Types de concepts	71
4.3.2	Attributs	72
4.3.3	Relations	74
4.3.4	Phrases	76
4.4	Discussion et conclusion	77
4.4.1	À propos de la méthode	77
4.4.2	À propos des résultats	78
4.4.3	Conclusion	79
5	Conception du langage graphique de visualisation des connaissances sur le médicament	80
5.1	Introduction et objectif	80
5.2	Matériel et méthodes	80
5.2.1	Analogie, conventions et symboles arbitraires	81
5.2.2	Grammaire combinatoire	82
5.2.3	Représentation graphique des relations <i>est-un</i>	84
5.2.4	Méthodes de combinaison de signes	85
5.2.5	Élaboration du lexique des pictogrammes	88
5.2.6	Formalisation de la grammaire du langage graphique	88
5.3	Résultats	89
5.3.1	Analyse des combinaisons de localisations et de troubles génériques	89
5.3.2	Représentation des attributs	89
5.3.3	Représentation des concepts	91
5.3.4	Représentation des phrases	91
5.3.5	Dessin des pictogrammes	94
5.3.6	Présentation du langage VCM version 1.0	94
5.3.6.1	Représentation des icônes	95
5.3.6.2	Représentation des phrases	96
5.3.6.3	Exemples	97
5.4	Discussion et conclusion	97
5.4.1	À propos de la méthode	97
5.4.2	À propos des résultats	97
5.4.3	Conclusion	98
6	Utilisation de VCM dans les textes médicaux et application de techniques de visualisation d'information	99
6.1	Introduction et objectifs	99
6.2	Matériel et méthodes	99
6.2.1	Matériel pour la réalisation du prototype	99

6.2.1.1	Langage de programmation	99
6.2.1.2	Données	99
6.2.2	Juxtaposition du texte et des icônes	100
6.2.3	Représentation à l'aide d'un schéma anatomique	100
6.2.4	Règles de combinaison d'icônes	102
6.2.5	Interaction médicamenteuse avec l'automédication	102
6.3	Résultats	103
6.3.1	Outil de lecture	103
6.3.2	"Monsieur VCM"	103
6.3.3	Outil de vérification	104
6.3.4	Outil d'aide à l'éducation du patient	106
6.3.5	Outil de suivi	108
6.4	Discussion et conclusion	109
6.4.1	À propos de la méthode	109
6.4.2	À propos des résultats	109
6.4.3	Conclusion	110
7	Évaluation du langage VCM	111
7.1	Introduction et objectifs	111
7.2	Matériel et méthodes	111
7.2.1	Recrutement des médecins évaluateurs	111
7.2.2	Évaluation 1 : compréhensibilité des icônes VCM	111
7.2.3	Évaluation 2 : comparaison de la vitesse de lecture avec les icônes VCM par rapport au texte	111
7.2.4	Évaluation 3 : comparaison de la vitesse de lecture avec "Monsieur VCM" par rapport au texte	112
7.2.5	Description des documents et des questions	112
7.2.6	Génération des RCP chimériques et des questions	112
7.2.7	Schéma des évaluations 2 et 3	114
7.2.8	Analyse statistique des évaluations 2 et 3	114
7.3	Résultats	115
7.3.1	Médecins évaluateurs	115
7.3.2	Matériel pédagogique	115
7.3.3	Évaluation 1 : compréhensibilité des icônes VCM	115
7.3.4	Caractéristiques des RCP chimériques générées	115
7.3.5	Évaluation 2 : comparaison de la vitesse de lecture avec les icônes VCM par rapport au texte	120
7.3.6	Évaluation 3 : comparaison de la vitesse de lecture avec "Monsieur VCM" par rapport au texte	120
7.4	Discussion et conclusion	120
7.4.1	À propos de la méthode	120
7.4.2	À propos des résultats	121
7.4.2.1	Temps d'apprentissage	121
7.4.2.2	Compréhensibilité des icônes	121
7.4.2.3	Temps de réponse	122
7.4.2.4	Taux d'erreur	122
7.4.2.5	Classification des erreurs	123
7.4.3	Conclusion	123

8	Discussion et conclusion	125
8.1	Comparaison avec les jeux de pictogrammes ou langages graphiques médicaux existants	125
8.2	Perspectives médicales	126
8.2.1	Amélioration et extension du langage VCM	126
8.2.2	Utilisation de VCM dans les logiciels et outils médicaux	126
8.2.3	Utilisation de VCM sur d'autres supports	126
8.2.4	Utilisation de VCM lors de la formation	127
8.2.5	Utilisation de VCM auprès d'autres professionnels de santé	127
8.2.6	Un langage graphique médical pour les patients	128
8.3	Perspectives en sciences cognitives	128
8.3.1	Une méthode rigoureuse de conception des langages graphiques	128
8.3.2	Validation des hypothèses	128
8.3.3	Un langage indépendant de la langue et de la culture	129
8.4	Perspectives en informatique et en représentation des connaissances	129
8.4.1	Traduction automatique de textes en VCM	129
8.4.2	Génération de traducteurs à partir de la grammaire graphique	130
8.4.3	Représentation graphique de terminologies ou d'ontologies	130
8.4.4	Visualisation et recherche d'information ou de connaissance?	131
8.5	Conclusion	132
A	Publications	144
B	Glossaire	145
C	Poésie en langage VCM	147
D	Rétrospective de la conception de VCM	148
E	Manuel d'apprentissage du langage VCM	150
E.1	Introduction	150
E.2	L'état actuel du patient	151
E.2.1	Les états physiologiques et le mode de vie du patient	151
E.2.2	Les pathologies et les symptômes	151
E.3	Les risques de pathologies ou de symptômes	153
E.4	Les antécédents de pathologies	154
E.5	Les traitements	154
E.5.1	Les types de traitement	154
E.5.2	Les médicaments	154
E.5.3	Les propriétés des traitements	155
E.6	Les surveillances	155
E.7	Construction de phrases	156
E.7.1	Combinaisons de plusieurs icônes en une seule	156
E.7.2	Négations	156
E.7.3	Actions	156
E.7.4	Relations ET et OU	157
E.7.5	Exemples	157
F	Dictionnaire VCM-français	159
	Dictionnaire VCM-français des états physiologiques et du mode de vie du patient	159
	Dictionnaire VCM-français des pathologies et symptômes	160
	Dictionnaire VCM-français des traitements	166
	Dictionnaire VCM-français des surveillances	171

G	Lexique des pictogrammes du langage VCM	173
	Lexique des pictogrammes et des formes	173
H	Description formelle du langage VCM en BNF	179
	H.1 Définition des connaissances exprimées par VCM	179
	H.2 Définitions des éléments graphiques utilisés par VCM	180
	H.3 Correspondance connaissances \Leftrightarrow éléments graphiques en VCM	180
	H.4 Exemple	182
I	Classification NPNM	183
J	Description du langage XML utilisé pour coder les connaissances médicales	184
	J.1 Présentation	184
	J.2 Exemple	185
	J.3 DTD	185

Table des figures

1.1	Raisonnement attendu devant le langage VCM.	17
1.2	Connaissance médicale, langage médical textuel et langage graphique.	17
2.1	La triade peircienne.	20
2.2	Exemples d'application des lois de la ségrégation de la <i>Gestalt</i>	21
2.3	Exemples et contre-exemples de variables sélectives, associatives, ordonnées et quantitatives.	22
2.4	Modèle de la vision	23
2.5	Exemple de perception pré-attentive.	24
2.6	Les deux niveaux d'interprétation et la conception d'une approche graphique.	24
2.7	Exemple d'hiéroglyphes : décomposition de l'idéogramme "sortir", d'après [1].	25
2.8	Quelques symboles de Sémantographie.	26
2.9	Exemples de panneaux de signalisation routière (d'après http://www.securiteroutiere.gouv.fr).	28
2.10	Structure sémiotique d'un panneau de danger d'après [97].	28
2.11	Pictogrammes de produits chimiques dangereux (d'après http://www.aimt67.org/symboles/etiquetage.htm).	29
2.12	Icônes + textes ou icônes seules dans le logiciel GAIM.	31
2.13	Exemples de signes médicaux extraits du "Symbol sourcebook" [42].	31
2.14	Exemple de pictogrammes de l'AFSSAPS.	32
2.15	Copie d'écran de Stabilis 2, d'après [155].	32
2.16	Exemples d'icônes issues de Stabilis 2, d'après [155].	33
2.17	Exemples de pictogrammes de l'USP, d'après http://www.usp.org	33
2.18	Représentation d'une pathologie avec UVAL-MED, d'après [117].	34
2.19	Exemple de repliement de code dans GNU Emacs (d'après http://mail.python.org/pipermail/python-list/2002-May/106656.html).	36
2.20	Lentille de document.	37
2.21	Mur fuyant représentant des fichiers triés par date.	37
2.22	Visualisation de code avec SeeSoft.	37
2.23	IPBC (<i>Interactive Parallel Bar Charts</i>) [30] représentant des données d'hémodialyse.	38
2.24	Visualisation de spectres antibactériens [44, 43].	38
2.25	Exemple de coordonnées parallèles, représentant des modèles de voitures disponible en applet sur http://www.cs.uta.fi/~hs/pce/	39
2.26	16 logiciels représentés par InfoBug.	40
2.27	Glyphe en étoile.	40
2.28	Lentille de tableau montrant la visualisation d'une base de données de joueurs de base-ball.	41
2.29	Copie d'écran de DOPAMINE.	42
2.30	Copie d'écran de SegusoLand.	42
2.31	Lentilles magiques dans un logiciel de dessin.	43
2.32	Exemples de diagrammes d'arbres.	44

2.33	Exemple d'arbre hyperbolique appliqué à la classification des espèces de primates (d'après http://www.inrp.fr/Acces/Biogeo/inxight/slsbeta/primates/primates.html).	44
2.34	Exemple d'arbre conique (d'après [152]).	45
2.35	Exemple de carte d'arbres appliquée à la représentation d'un système de fichiers (copie d'écran de http://treemap.sf.net).	46
2.36	Représentation des arbres parallèles en taxonomie botanique par M.J. Graham [62, 61].	46
2.37	Exemple de <i>Fisheye</i> déformant, d'après [132].	47
2.38	Terrain 3D généré par VxInsight [16] à partir d'un ensemble de textes de brevets . . .	48
2.39	Exemples d'approches graphiques basées sur des forces d'attraction et de répulsion. .	48
2.40	VIBE [104] (d'après [68]).	49
2.41	Exemple d'arbre botanique (d'après [80]).	50
2.42	Les 4 formes de la connaissance : textuelle, structurée, graphique et graphique interactive.	53
3.1	Modèle UML des indications des RCP, d'après [43, 45, 154]	56
3.2	Modèle UML des contre-indications des RCP, d'après [93]	57
3.3	Modèle UML des interactions médicamenteuses des RCP, d'après [43]	58
3.4	Modèle UML des effets indésirables des RCP, d'après [43]	58
3.5	Modèle UML de la pharmacocinétique, version simplifiée d'après [43].	59
3.6	Copie d'écran de la monographie de l'amiodarone dans le Vidal électronique.	59
3.7	Copie d'écran de la monographie de l'amiodarone dans la Banque Claude Bernard. . .	60
3.8	Exemple d'arbre de décision extrait d'un guide de bonnes pratiques sur l'angine de poitrine (d'après http://www.openclinical.org/).	61
4.1	Méthode utilisée pour l'analyse sémantique des concepts médicaux.	69
4.2	Exemple d'utilisation de la classification ATC pour enrichir les dénominations des médicaments.	70
4.3	Méthode utilisée pour l'analyse de la structure des phrases.	71
4.4	Modèle UML des états du patient (à gauche) et des phrases médicales (à droite). . . .	75
4.5	Exemples d'utilisation de notre modèle.	77
5.1	Exemples de représentation de l'angor (ou angine de poitrine) sans ou avec une grammaire combinatoire.	82
5.2	Exemples d'utilisation de la théorie de la <i>Gestalt</i> pour déterminer les relations <i>est-un</i> graphiques.	84
5.3	Exemples de relations <i>est-un</i> graphiques.	85
5.4	Décomposition des différents concepts et attributs en arbre (issu du modèle UML de la figure 4.4).	87
5.5	Exemples de pictogramme à forme creuse (à gauche) et à forme pleine (à droite). . . .	87
5.6	Exemple d'application de notre formalisme à la grammaire des panneaux routiers (version simplifiée).	88
5.7	Décomposition des différents concepts et attributs en arbre (issu du modèle UML de la figure 4.4).	91
5.8	Représentation formelle en UML de la grammaire et de la sémantique des icônes du langage VCM.	92
5.9	Représentation formelle en UML de la grammaire et de la sémantique des phrases du langage VCM.	93
5.10	Construction des icônes du langage VCM.	94
6.1	Exemples de schémas anatomiques.	101
6.2	Exemples de combinaisons d'icônes.	102
6.3	Copie d'écran de l'outil de lecture, pour l'Arkogélule Millepertuis (un antidépresseur). .	103
6.4	La répartition des différentes localisations et étiologies sur "Monsieur VCM".	104

6.5	Copie d'écran de l'outil de vérification, pour le Stilnox (un somnifère).	105
6.6	Copie d'écran de l'outil de vérification, pour le Stilnox, après avoir cliqué sur l'icône des contre-indications en cas de troubles musculaires.	105
6.7	Copie d'écran de l'outil de vérification, pour le Stilnox, après avoir cliqué sur l'icône des précautions d'emploi en cas de troubles psycho-comportementaux.	106
6.8	Copie d'écran de l'outil d'aide à l'éducation du patient, pour le Stilnox (un somnifère).	107
6.9	Copie d'écran de l'outil d'aide à l'éducation du patient pour le Stilnox, après avoir cliqué sur l'icône des effets indésirables oculaires.	107
6.10	Copie d'écran de l'outil d'aide au suivi pour le Stilnox.	109
7.1	Méthode utilisée pour générer les RCP chimériques et les questions à partir de la base Thériaque, et pour construire les séquences.	113
7.2	Copies d'écran du didacticiel VCM.	116
7.3	Extrait de l'évaluation de compréhensibilité.	117
7.4	Résultat de l'évaluation de compréhensibilité.	117
7.5	Copies d'écran de l'évaluation de vitesse.	118
7.6	Résultats de l'évaluation de la vitesse de lecture de VCM.	118
7.7	Copies d'écran de l'évaluation de "Monsieur VCM".	119
7.8	Résultats de l'évaluation de "Monsieur VCM".	119
8.1	Exemple de représentation d'une partie d'une terminologie sous forme de réseau.	130

Liste des tableaux

2.1	Les différentes variables rétinienne et leurs propriétés d'après J. Bertin.	22
2.2	Caractéristiques des différentes techniques de visualisation d'information.	51
2.3	Les différents usages des approches graphiques.	52
2.4	Exemples d'applications des différents usages, pour le patient et pour les professionnels de santé.	52
4.1	Les attributs, leurs valeurs et leurs fréquences dans 5 types de concept.	73
5.1	Avantages et inconvénients des analogies, conventions et symboles arbitraires.	81
5.2	Propriétés des différentes méthodes de combinaison de signes.	86
5.3	Nombre de valeurs et variables rétinienne possibles pour chaque attribut.	90
5.4	Analogies et conventions existantes pour représenter les notions présentes dans les phrases VCM.	92
7.1	Caractéristiques moyennes des documents de l'évaluation 2.	117
7.2	Caractéristiques moyennes des documents de l'évaluation 3.	117

Chapitre 1

Introduction

1.1 Problématique

Les connaissances médicales sont de plus en plus nombreuses et de plus en plus complexes. La découverte de nouvelles pathologies et de nouveaux traitements sont responsables de l'augmentation du volume et de la complexité des connaissances médicales disponibles. Cela est vrai en particulier pour les connaissances sur le médicament, comme les monographies des médicaments ou les Résumés des Caractéristiques Produit (RCP).

Ces connaissances médicales sont jugées utiles par les professionnels de santé [67], cependant le volume important des connaissances rend leur utilisation difficile en pratique clinique, notamment lorsque le professionnel est en face du patient et dispose donc d'un temps très limité. Par exemple, quel médecin a le temps, en consultation, de lire les 10 RCP correspondant aux 10 médicaments qui figurent sur l'ordonnance d'un patient polymédiqué? J.W. Ely *et al.* ont montré que **le temps est le premier obstacle que le médecin rencontre lorsqu'il recherche des connaissances médicales** [54], et qu'en consultation, le médecin alloue une durée fixe, en moyenne inférieure à deux minutes, pour chacune de ces recherches; lorsque cette durée est écoulée, il abandonne la recherche [53].

Ce problème est bien connu et des solutions ont été proposées. Les connaissances médicales (dont les RCP) sont maintenant mises à disposition dans des formats électroniques (CD-ROM, Internet), et des systèmes d'aide à la décision ont été proposés. Une première approche proposée est l'approche critique, dans laquelle le système critique la prescription du médecin. Ces systèmes agissent en déclenchant des alertes [57] lorsque la prescription du médecin n'est pas en conformité avec les connaissances dont dispose le système, et éventuellement en proposant des améliorations [94].

Une autre approche est l'approche guidée; celle-ci a été proposée notamment par B. Seroussi *et al.* dans ONCODOC [135]. Ce système permet au médecin de naviguer dans un arbre de décision qui correspond par exemple à un Guide de Bonne Pratique (GBP), *via* un système de questions-réponses. Le médecin participe alors de manière active au raisonnement. L'approche critique et l'approche guidée sont complémentaires, et certains projets comme ASTI (Aide à la Stratégie Thérapeutique Informatisée, [136, 47, 48]) proposent les deux.

Si ces approches sont intéressantes et ont obtenu des résultats positifs, elles nécessitent en revanche une phase de développement et une maintenance importante, afin par exemple d'informatiser les Guides de Bonne Pratique, puis de les relier à un moteur d'inférence et de mettre régulièrement à jour les connaissances. De plus, elles peuvent avoir des conséquences négatives, la plus fréquente étant de demander du travail supplémentaire au médecin, par exemple pour le codage des pathologies du patient [22]. Par ailleurs, les deux approches présentées ne s'appliquent pas à l'ensemble des connaissances médicales: par exemple, présenter au médecin les effets indésirables d'un médicament sous forme d'alertes ou d'arbre de décision apparaît inapproprié.

D'après une enquête sur les bases de connaissances sur le médicament réalisée par FULMEDICO (Fédération des Utilisateurs de Logiciel Médicaux et Communicants) [58], "on a [...] l'impression d'être en présence d'une informatisation des bases de données médicamenteuses 'papier' ". Les possibilités

aujourd'hui offertes par le multimédia ne sont pas, ou insuffisamment, mises à profit.

De nombreuses études [13, 109, 74, 123] ont mis en évidence **les capacités de la vision humaine en terme d'analyse d'images** ; ces capacités sont sous-utilisées par la présentation textuelle des connaissances. Dans d'autres domaines, comme la signalisation routière (panneaux routiers), la représentation du danger (étiquetage des produits chimiques) ou la modélisation (langage UML, *Unified Modeling Language*), l'utilisation d'approches graphiques a permis d'accélérer l'accès aux connaissances. L'étude bibliographique que nous présenterons au chapitre 2 a fait ressortir deux grands types d'approches graphiques complémentaires :

- **les langages graphiques**, qui permettent de représenter des informations ou des connaissances en combinant des pictogrammes, des couleurs, des formes,..., selon une grammaire précise (par exemple les panneaux routiers),
- **la visualisation d'information**, qui propose une grande variété de techniques pour représenter graphiquement la structure d'éléments d'information (par exemple la représentation de structures arborescentes).

Il semble raisonnable de penser que la médecine pourrait, elle aussi, profiter de ces approches. Plusieurs travaux ont déjà été réalisés pour visualiser des données médicales [29, 51], par exemple des résultats d'analyse. Cependant, très peu d'approches graphiques ont été proposées pour la visualisation des connaissances médicales, et la quasi-totalité de celles-ci sont destinés aux patients [37, 73], et non aux professionnels de santé. Ces approches ont alors pour objectif d'attirer l'attention du patient, de rendre les connaissances accessibles aux patients ne pouvant pas lire les textes, ou d'expliquer au patient le mécanisme du processus pathologique ou de son traitement. En revanche, elles ne permettent pas d'accélérer l'accès aux connaissances pour le médecin.

Dans cette thèse, nous nous sommes placé dans un cadre multi-disciplinaire : médecine et pharmacie, sciences cognitives et informatique, pour mettre au point des méthodes de visualisation des connaissances médicales, dans le but de faciliter et d'accélérer l'accès aux connaissances par les professionnels de santé. Ces méthodes comprendront **un langage graphique de Visualisation des Connaissances Médicales (VCM)** ainsi que des techniques de visualisation d'information adaptées à ce langage graphique et aux textes médicaux.

La plupart des approches graphiques existant actuellement, dans le domaine médical ou en-dehors, ont été développées de façon empirique et souvent sans véritable évaluation. Nous avons néanmoins souhaité utiliser une méthodologie rigoureuse prenant en compte à la fois la nature des connaissances sur le médicament et les capacités de la vision humaine, ainsi qu'une évaluation rigoureuse effectuée dans des conditions contrôlées.

1.2 Objectifs

Notre objectif est :

- de mettre au point un langage graphique de Visualisation des Connaissances Médicales (VCM) facile à apprendre pour les professionnels de santé, et qui leur facilite et accélère l'accès aux connaissances médicales lors des consultations,
- de mettre au point des techniques de visualisation d'information adaptées à ce langage graphique et aux textes médicaux,
- d'évaluer ce langage VCM et ces techniques de visualisation d'information de manière rigoureuse.

Dans cette thèse, **nous nous sommes focalisé sur les connaissances sur le médicament**, et plus particulièrement les Résumés des Caractéristiques Produit (RCP) des spécialités médicamenteuses (les connaissances disponibles sur d'autres types de médicaments, comme l'homéopathie ou l'herboristerie, ne sont pas forcément de même nature et peuvent poser des problèmes différents). Les connaissances sur le médicament représentent une part importante des connaissances médicales, et font notamment référence à des pathologies, des traitements, des examens,...

Les principales questions sur le médicament que se posent les professionnels en situation clinique, et auxquelles VCM devra permettre de répondre, sont les suivantes :

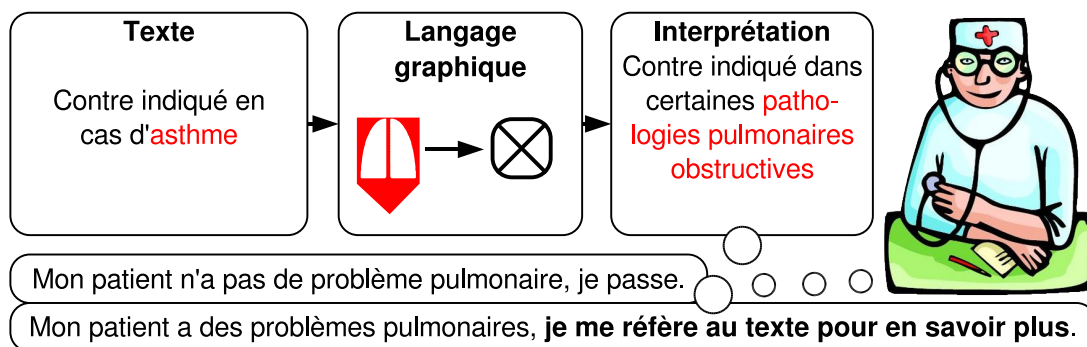


Fig. 1.1 – Raisonement attendu devant le langage VCM.

Le passage par le langage graphique implique une perte de précision (ici, l’asthme devient une pathologie pulmonaire obstructive).

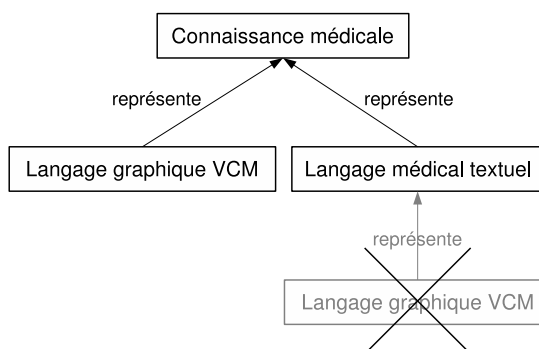


Fig. 1.2 – Connaissance médicale, langage médical textuel et langage graphique.

- quelles sont les principales propriétés du médicament X ?
- le médicament X est-il contre-indiqué chez mon patient Y ?
- que dois-je dire sur le médicament X à mon patient Y ?

Le langage VCM n’a pas pour objectif d’être aussi précis que les textes médicaux, ni de remplacer ceux-ci. Au contraire, l’objectif est d’enrichir les textes existants avec des icônes, ou de servir de “filtre” afin de permettre au médecin d’éliminer les passages du texte qui ne concernent pas son patient, et de ne lire que les passages intéressants pour ce patient (voir figure 1.1). Dans tous les cas, le texte reste la référence.

Les langages graphiques peuvent être indépendants de la langue et / ou de la culture. Nous souhaitons que le langage VCM soit aussi indépendant de la langue et de la culture françaises que possible, cependant cela n’est pas notre objectif premier. En effet, pour mettre au point un langage graphique culturellement indépendant, il nous aurait fallu une équipe multi-culturelle, ce dont nous ne disposons pas.

1.2.1 VCM et le langage naturel

Dans cette section, nous souhaitons préciser le rapport entre le langage VCM et le langage naturel. Le langage VCM a pour objectif de représenter graphiquement les connaissances médicales. En revanche, l’objectif de VCM **n’est pas** de représenter le langage médical textuel utilisé d’ordinaire pour représenter les connaissances médicales (voir figure 1.2). Il ne s’agit donc pas de traduire mot-à-mot le français médical en icônes, par exemple de traduire “médicament de l’asthme” par trois icônes signifiant “médicament”, “de” (possessif) et “asthme”. En particulier, le langage VCM doit être indépendant des tournures utilisées dans les textes, des raccourcis et des abréviations.

1.3 Plan de la thèse

Tout d'abord, nous présenterons l'état de l'art des modèles de la vision, des langages graphiques et de la visualisation d'information, et nous conclurons sur la démarche à suivre pour la conception de VCM (chapitre 2). Nous verrons ensuite l'état de l'art des connaissances sur le médicament et des besoins exprimés par les différents acteurs en la matière, et nous proposerons une classification des outils de visualisation des connaissances sur le médicament (chapitre 3).

Afin de concevoir le langage VCM, nous montrerons dans un premier temps comment nous avons construit un modèle des connaissances sur le médicament adapté, à partir de connaissances d'experts, de modèles de la littérature et de l'analyse d'un corpus de RCP (chapitre 4). Nous verrons ensuite comment nous avons défini la grammaire du langage VCM en nous appuyant sur des règles mises au point à partir des sciences cognitives (chapitre 5), puis des méthodes de visualisation d'information permettant de combiner texte et icônes VCM (chapitre 6).

Enfin, nous présenterons l'évaluation dans des conditions contrôlées d'un prototype de base de connaissances sur le médicament utilisant le langage VCM, sur un groupe de médecins formés à VCM (chapitre 7). Nous terminerons en discutant les perspectives ouvertes par ce travail (chapitre 8).

Un glossaire est disponible en annexe B, un manuel du langage VCM en annexe E et F, et un lexique des pictogrammes de VCM en annexes G.

Chapitre 2

État de l'art des modèles de la vision, des langages graphiques et de la visualisation d'information

Au cours de l'étude bibliographique, nous avons constaté que les approches graphiques présentaient deux avantages sur les approches traditionnelles à base de texte :

L' "**universalité**" la phase d'apprentissage est courte, voire inexistante, et les approches graphiques peuvent être quasi-indépendantes de la langue ou de la culture. En effet, l'utilisation du principe d'analogie permet de représenter des concepts de façon quasi-universelle, par exemple il est facile de deviner que le dessin d'un oeil signifie "oeil". Cependant, l'analogie est incapable de représenter certaines notions abstraites, comme l'interrogation ou la négation. Il existe aussi un certain nombre de conventions graphiques très répandues, comme le symbole "coeur" utilisé par les cartes à jouer.

La rapidité de lecture en tirant parti des possibilités visuelles, les représentations graphiques peuvent être lues beaucoup plus rapidement, tout en étant plus "attractives".

Cependant, ces approches graphiques ont deux inconvénients :

Des moyens technologiques plus complexes sont nécessaires : impression en couleur, animation,...

La précision est souvent inférieure à celle du texte même s'il existe des exceptions, par exemple les cartes géographiques sont plus précises qu'un texte équivalent.

En fonction de ces avantages et inconvénients, les approches graphiques seront plus ou moins bien adaptées aux différentes connaissances médicales et à leur utilisation en pratique.

Une étude bibliographique nous a permis de mettre en évidence deux types d'approches graphiques complémentaires : **les langages graphiques et la visualisation d'information**. Un langage graphique permet de traduire graphiquement des informations ou des connaissances, tandis que la visualisation d'information représente de manière graphique la structure de l'information, et ajoute souvent de l'interactivité.

2.1 La sémiotique et les modèles de la vision

La *sémiotique* est l'étude des signes et des systèmes de signes ; contrairement à la sémantique, elle ne se limite pas aux langages textuels et essaie d'étendre les projets de la linguistique à d'autres systèmes de signes comme les langages graphiques [151]. La plupart des travaux de sémiotique s'intéressent à la classification et l'analyse de langages existants. Quelques travaux s'appliquent cependant à la conception de langages graphiques ; nous allons en présenter brièvement quelques-uns, en nous focalisant sur les aspects les plus pratiques.

Ces travaux s'inspirent de modèles psychologiques ou neurophysiologiques de la perception visuelle.

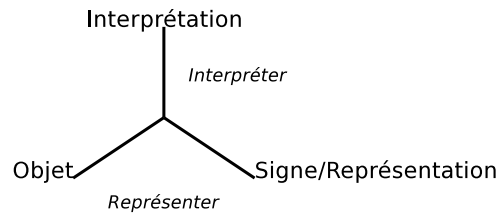


Fig. 2.1 – La triade peircienne.

2.1.1 Sémiotique

Selon Peirce, un *signe* ou Représentation R (*representamen*) représente un Objet O qui est interprété en une Interprétation I : c'est la "triade peircienne" (figure 2.1, [50, 8]). La relation objet-signe est la relation de représentation, la relation signe-interprétation est la relation d'interprétation. Ces deux relations sont nécessaires pour une bonne compréhension de l'objet à travers le signe.

Peirce distingue trois types de signes :

Le signe iconique ressemble à l'objet qu'il représente (principe d'analogie, le signe iconique est parfois appelé "icône", mais le sens est différent de celui du mot "icône" en informatique ; aussi nous parlerons de signe iconique pour éviter la confusion).

L'index est un signe associé à l'objet qu'il représente (par exemple de la fumée pour du feu).

Le symbole est un signe dont la signification est arbitraire et conventionnelle, sans rapport avec l'objet qu'il représente (par exemple les mots, cf F. de Saussure).

Ces 3 catégories sont loin d'être exclusives, et un même signe peut appartenir à l'une ou l'autre selon le sens qui lui est associé : par exemple le dessin d'une imprimante est un signe iconique s'il signifie "imprimante", un index s'il signifie "impression" et un symbole s'il signifie "enregistrer" (ce dernier cas était, bien sûr, à éviter!).

Les signes iconiques peuvent être compris sans apprentissage préalable, en procédant par *analogie* (voir <http://www.edusud.org/ressources/ntic/ica/index.html>). Ce n'est pas le cas des symboles, cependant les abstractions suivantes ne peuvent être représentées *que* par des symboles : la négation, l'interrogation, la généralisation, les relations temporelles (avant, après, pendant,...), les relations de cause et de conséquence,...

Selon la théorie du *double codage* [107], les mécanismes cognitifs mis en jeu lors de la perception sont distincts selon le type de signe. Il y aurait deux modes de représentation mentale distincts : un système de représentations arbitraires, verbales ou propositionnelles correspondant au traitement des symboles (langage, abstraction,...), et un second système de représentations figuratives basé sur une "sémantique de la ressemblance" ou analogie, correspondant aux signes iconiques.

2.1.2 Psychologie de la forme (théorie de la *Gestalt*)

La psychologie de la forme ou théorie de la *Gestalt* ([109, 7, 159], http://fr.wikipedia.org/wiki/Psychologie_de_la_forme) a été fondée aux débuts des années 1920 par des psychologues allemands. Une forme *Gestalt* est un ensemble structuré et doté de sens (*Gestalt* en allemand signifie forme mais aussi structure, organisation ; nous parlerons de "forme *Gestalt*" pour éviter toute confusion avec le sens commun de "forme"). Les principes de base de la psychologie de la forme sont les suivants :

- **Un ensemble est plus que la somme de ses parties** [159]. Par exemple une image est composée de points colorés (des pixels sur un écran, des taches lumineuses sur la rétine,...) mais contient plus que ces points : la position de chaque point *par rapport à l'ensemble des autres points* doit aussi être prise en compte ; ce sont ces positions relatives qui constituent la forme. En conséquence, une forme *Gestalt* ne peut pas être décomposée en sensations indépendantes.
- **Les formes *Gestalt* sont pré-existantes** : elles existent en dehors de l'esprit qui la perçoit, et sont donc au moins partiellement indépendantes de l'observateur (ce qui est contraire à

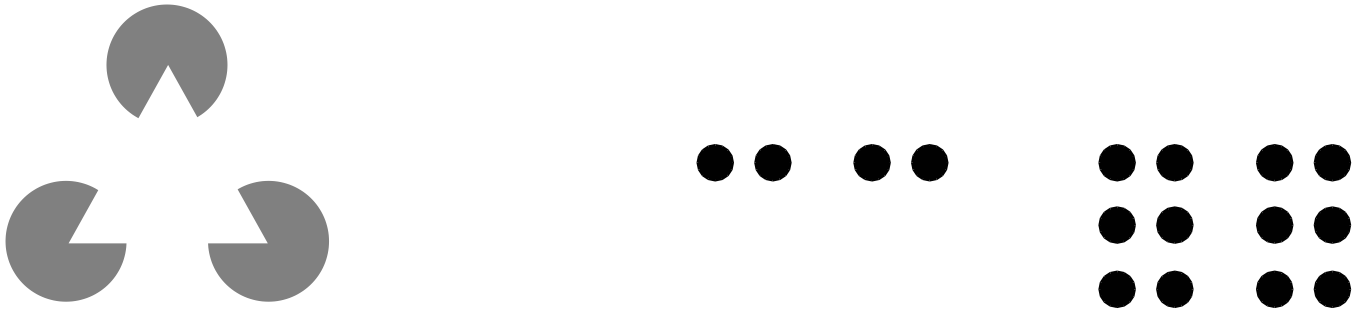


Fig. 2.2 – Exemples d’application des lois de la ségrégation de la *Gestalt*.

A gauche : exemple d’application de la loi de clôture : on voit un triangle blanc, bien que celui-ci ne soit pas complet. A droite : exemple d’application de la loi de répétition : il est plus facile de distinguer les deux colonnes de points à droite, à cause de leur répétition sur trois lignes.

d’autres théories comme l’empirisme). Ce principe découle en fait du précédent, par exemple si ce sont les positions de chaque point les uns par rapport aux autres qui constituent la forme *Gestalt* d’une image, celles-ci ne dépendent pas de l’observateur.

Cette théorie générale a notamment été appliquée à la compréhension des mécanismes de la perception visuelle : Pourquoi dans une image donnée percevons-nous telle forme plutôt qu’une autre ? Par exemple, dans la figure 2.2 à gauche, voyons-nous un triangle et trois cercles, ou bien trois cercles dont il manque un segment ? Très souvent, plusieurs formes peuvent être perçues dans une image, cependant l’une d’entre elle est la *bonne forme*, celle qui sera perçue par l’œil. La bonne forme est indépendante de la personne qui regarde l’image. Plusieurs lois permettent de déterminer quelle est la bonne forme :

- **La loi de la bonne forme ou prégnance** : la forme reconnue est la plus simple, régulière et symétrique parmi toutes celles possibles.
- **Les lois de ségrégation des unités** :
 - La loi de proximité : les éléments proches tendent à être regroupés dans une même forme (par exemple les étoiles en constellation).
 - La loi de ressemblance : les éléments semblables (même couleur, même taille, ...) tendent à être regroupés dans une même forme.
 - La loi de symétrie : les formes ayant un axe de symétrie sont mieux perçues que les autres.
 - La loi de clôture : les formes tendent à être “complées” ou “complétées” pour se rapprocher des formes connues très prégnantes (voir figure 2.2 à gauche).
 - La loi de répétition : lorsqu’une loi se répète plusieurs fois, son effet augmente (voir figure 2.2 à droite).

En tenant compte de ces lois, il est possible de déterminer ce qui est perçu dans une image complexe, comprenant de multiples éléments.

2.1.3 Sémiologie graphique

La *sémiologie graphique* est issue des travaux du cartographe J. Bertin [12]. Elle définit les caractéristiques et les composantes d’un graphique et leur impact cognitif. Un graphique est composé de taches en implantation ponctuelle (les taches sont des points), linéaire (des lignes) ou zonale (des surfaces). Chaque tache a trois dimensions : x et y , coordonnées spatiales dans le plan, et z , qui correspond à une ou plusieurs variables rétinienne(s) (voir la liste dans le tableau 2.1).

D’autres auteurs (M. Green notamment) ont ajouté de nouvelles variables : le mouvement (qui se décompose en deux sous-variables : la direction et la vitesse), le clignotement (deux sous-variables : la fréquence et la phase) et la disparité (affichage d’images différentes pour l’œil droit et l’œil gauche, avec des lunettes de réalité virtuelle par exemple). Cependant, le clignotement attire voire détourne l’attention de l’utilisateur¹, ce qui peut être très désagréable.

¹Voir par exemple le langage HTML de mise en forme de page web, où la balise de clignotement `<blink>` n’est

variable	exemple	sélective	associative	ordonnée	quantitative	nombre de valeurs
x, y		++	+	+	+	élevé
taille		+		+	+	5
valeur		+		+	(+)	3
saturation		+		+	(+)	3
grain		+	+	+		3
couleur		++	+	(+)		10
orientation		+	+			12
forme		(+)	+	(+)	(+)	quasi-infinie

Tab. 2.1 – Les différentes variables rétinienne et leurs propriétés d’après J. Bertin.

Pour les propriétés, “+” indique que la propriété est présente pour une variable, “++” qu’elle est très prononcée, “(+)” qu’elle peut être présente mais à certaines conditions (changement d’échelle,...).

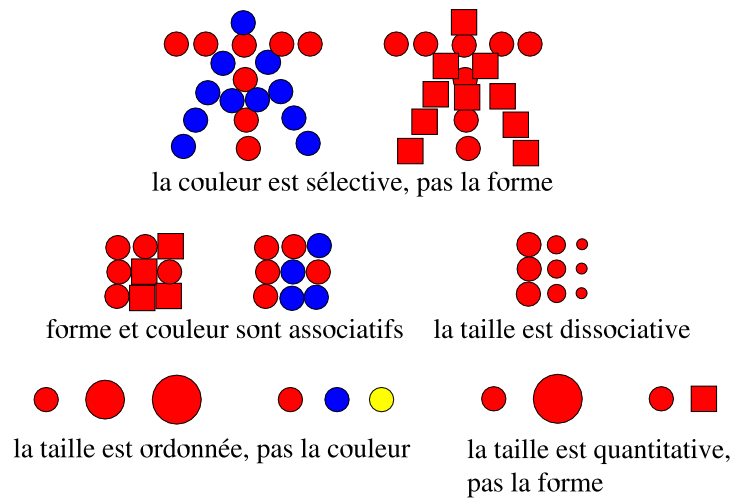


Fig. 2.3 – Exemples et contre-exemples de variables sélectives, associatives, ordonnées et quantitatives.

C. Bruley [18] a réalisé une présentation détaillée de la sémiologie graphique et des aspects cognitifs sous-jacents.

Quatre propriétés différencient ces variables ; ces propriétés sont appelées *niveaux d’organisation* par Bertin (voir exemple figure 2.3). Une variable peut être :

Sélective il est possible de sélectionner visuellement de façon immédiate les éléments pour lesquels la variable a une valeur donnée, et de faire abstraction des autres éléments.

Associative il est visuellement possible de faire abstraction des différentes valeurs de cette variable.

Ordonnée les différentes valeurs de la variable ont un ordre évident.

Quantitative il est visuellement possible de traduire la différence entre deux valeurs par un rapport numérique.

Le tableau 2.1 récapitule à la fois les niveaux d’organisation et les autres propriétés de chaque variable. Une autre propriété intéressante est le nombre de valeurs différentes utilisables pour une variable dans un même graphique. Les variables rétinienne ne sont clairement pas toutes égales, et certaines sont plus intéressantes, notamment x et y qui sont les seules à posséder les 4 niveaux d’organisation, la forme qui possède un nombre de valeurs quasi-infini, et la couleur qui a une très bonne sélectivité. Lors de la réalisation d’un graphique, il faut faire correspondre à chaque attribut des objets du graphique une variable rétinienne ; le choix de cette variable doit se faire en fonction presque plus utilisée.

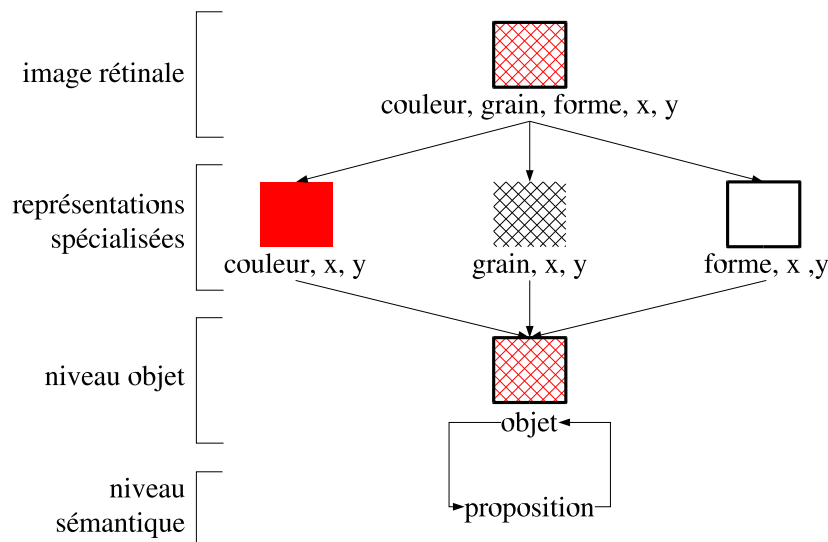


Fig. 2.4 – Modèle de la vision

des propriétés de la variable et de l'attribut. Par exemple si l'attribut est une valeur quantitative, nous choisirons une variable quantitative pour le représenter : x , y ou taille.

La sémiologie graphique était d'abord **une approche empirique** ; elle a ensuite été reliée à des modèles neurophysiologiques [63], comme celui présenté à la section suivante.

2.1.4 Modèle neurophysiologique

La figure 2.4 présente un modèle de la vision sur lequel un certain nombre d'auteurs sont d'accord ([63, 122], [74] chapitre 6 paragraphe *units of function in the cortex*). Ce modèle présente quatre niveaux :

l'image rétinienne correspond à une image "bitmap", c'est-à-dire un ensemble de points lumineux ordonnés les uns par rapport aux autres.

les représentations spécialisées correspondent chacune à un "module indépendant" prenant en charge une des propriétés de l'objet ; notons que la position (x, y) est présente dans tous les modules, et permet de réunir les représentations spécialisées par la suite,

le niveau objet permet de distinguer les objets et correspond à une image "vectorielle", c'est-à-dire un ensemble d'objets géométriques (carré, cercle,...) ayant divers attributs (position dans l'espace, couleur,...),

le niveau propositionnel ou sémantique cherche à "donner un sens" aux objets.

Ce modèle explique la prédominance des variables x et y constatée empiriquement par J. Bertin : en effet ce sont les seules variables présentes dans toutes les représentations spécialisées, alors qu'il n'existe pas de représentation spécialisée associant par exemple la couleur et la forme.

Les premières étapes de la perception semblent indépendantes de la volonté, des connaissances ou des attentes (elles sont impénétrables aux fonctions cognitives [122], à rapprocher de la notion de pré-existence des formes dans la théorie de la *Gestalt*), contrairement aux dernières étapes (niveau objet sur la figure 2.4) qui interagissent avec le niveau sémantique. Le sens donné à un objet peut alors modifier la perception de celui-ci.

2.1.5 Perception pré-attentive

On parle de *perception pré-attentive* lorsqu'un signe ou un groupe de signes semblent "surgir" hors de l'image, et peuvent donc être distingués très rapidement (moins de 200 ms) ; par opposition à la *perception attentive*, où l'utilisateur doit passer en revue chacun des signes un par un pour trouver ceux qui l'intéressent, et où le temps nécessaire est proportionnel au nombre de signes.

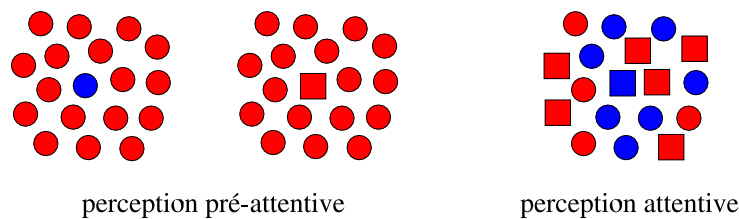


Fig. 2.5 – Exemple de perception pré-attentive.

Sur les deux images de gauche, la présence d'un unique rond bleu ou d'un unique carré se distingue immédiatement. A droite, il est nécessaire de parcourir l'ensemble des points pour s'apercevoir qu'il n'y a qu'un seul carré bleu.

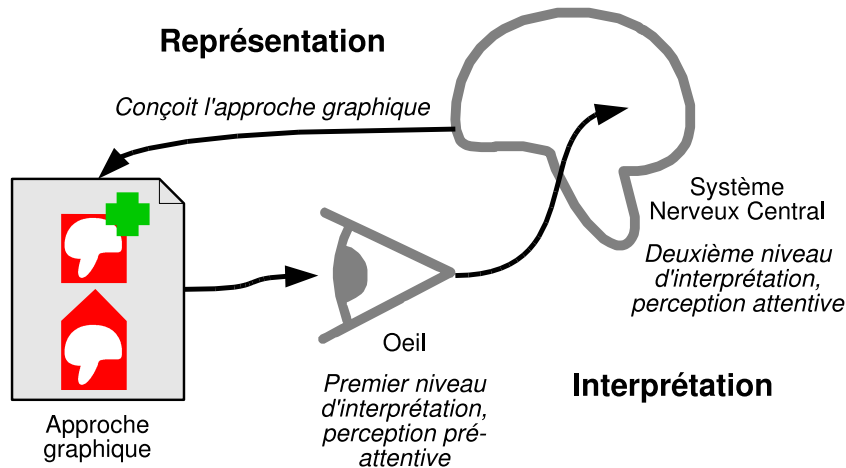


Fig. 2.6 – Les deux niveaux d'interprétation et la conception d'une approche graphique.

Il a été montré que la perception pré-attentive est possible lorsque la recherche porte sur **au plus une** variable rétinienne [160, 121] (voir exemple figure 2.5). Si deux variables ou plus sont en jeu, il s'agit d'une recherche conjonctive qui nécessite la perception attentive. Le modèle de perception visuelle présenté ci-dessus explique cela : la perception pré-attentive a lieu au niveau des représentations spécialisées (où une seule variable rétinienne est disponible, une par module indépendant), alors que la perception attentive a lieu aux niveaux supérieurs (objet et propositionnel).

Z. Pylyshyn [123, 122, 120, 121] présente un autre aspect de la vision pré-attentive. Quelques (3-6? 5-9?) objets sont indexés ('*FINSTed*') par la vision et peuvent être accédés ou manipulés de manière pré-attentive ; ils possèdent la propriété de sélection définie par Bertin. Il n'y a aucune contrainte sur la nature de ces objets (en particulier, ils n'ont pas besoin d'être proches les uns des autres ni similaires, et ils peuvent même être identiques à d'autres objets non indexés). Les objets indexés sont indépendants et peuvent se déplacer ; ils seront 'suivis'.

Par exemple, il est possible de déterminer le nombre d'objets sans avoir à les compter un à un lorsque ce nombre ne dépasse pas 3-6. Nous pouvons donc porter notre attention sur un petit nombre d'objets et pas uniquement sur un seul, comme cela est communément admis.

2.1.6 Conclusion sur les modèles de la vision

Dans cette section, nous avons présenté plusieurs modèles de la vision humaine. Ces modèles font souvent des hypothèses contradictoires : en particulier, la théorie de la *Gestalt* considère qu'il n'est pas possible de diviser une forme *Gestalt* en plusieurs sensations indépendantes, alors que la sémiologie graphique décompose l'image en taches et en couleurs, formes,..., considérées comme indépendantes. Pour autant, cela ne signifie pas nécessairement qu'un de ces modèles soit "faux", puisque chacun présente des limites. La réalité est certainement plus complexe : les formes perçues peuvent être décomposées en sensations indépendantes seulement dans certaines situations, lorsqu'il n'y a pas d'interaction entre elles.

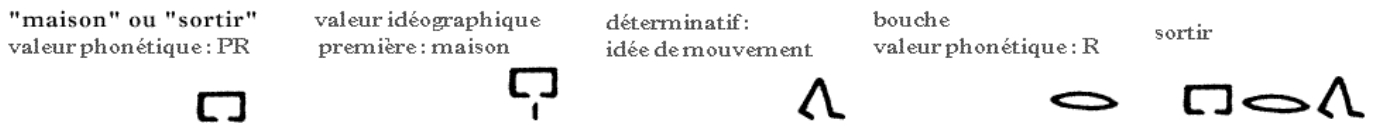


Fig. 2.7 – Exemple d’hiéroglyphes : décomposition de l’idéogramme “sortir”, d’après [1].

Notons le mélange d’élément graphique et phonétique : le signe “maison” suivi du signe “bouche” signifie que la prononciation est celle du signe maison (son “PR”).

Nous présentons ici un modèle simplifié de la vision sur deux niveaux (figure 2.6). L’interprétation d’un aspect d’un signe graphique peut se faire soit au premier niveau, soit au second, selon la nature de cet aspect. Le premier niveau d’interprétation est un niveau rapide, voire immédiat, et peut traiter un nombre quasi-infini de signes en parallèle. Ce niveau est quasi-indépendant de l’observateur et ne peut pas accéder aux fonctions cognitives (raisonnement), ou seulement de manière très limitée (mémoire). Nous avons appelé “oeil” ce premier niveau d’interprétation, même si celui-ci ne se déroule pas forcément dans l’oeil (le nerf optique étant considéré comme partie intégrante du Système Nerveux Central (SNC) en physiologie). Le second niveau ne peut traiter qu’un petit nombre de signes en parallèle ; si le nombre de signes est trop important, il faudra les traiter les uns après les autres. Il interagit directement avec les fonctions cognitives comme la mémoire ou le raisonnement.

Pour être lues rapidement, et de manière identique quel que soit le lecteur, **les approches graphiques doivent utiliser au maximum le premier niveau d’interprétation**, en particulier lorsque le nombre de signes présents simultanément est important. En revanche, la conception d’une approche graphique ne peut pas être le fait de l’oeil, et ne peut être réalisée que par les fonctions cognitives du SNC, qui doit “se mettre à la place” de l’oeil. Concevoir une approche graphique intuitive n’est donc pas une tâche intuitive, la notion d’“intuitif” n’étant pas la même au niveau de l’oeil et du SNC.

2.2 Les langages graphiques

Un langage graphique est un langage permettant d’exprimer des informations ou des connaissances sous forme graphique, à l’aide d’une combinaison de pictogrammes, de couleurs, d’icônes,..., et non textuelle [25]. Les différents éléments graphiques, pictogrammes, couleurs,..., sont combinés selon des règles précises, correspondant à une grammaire graphique [98].

Aujourd’hui les langages graphiques sont partout autour de nous, de la signalisation routière aux pictogrammes présents dans les gares en passant par les schémas utilisés dans l’éducation et par le langage de modélisation UML (*Universal Modeling Language*, <http://uml.org>). Dans cette section, nous présentons quelques exemples de langages graphiques. Ceux-ci étant trop nombreux pour en réaliser une liste exhaustive, nous avons plutôt sélectionné ceux qui nous paraissent les plus marquants ou intéressants.

2.2.1 Écritures graphiques

Un *logogramme* est un signe écrit représentant un mot. Il existe plusieurs types de logogramme : le *pictogramme* est figuratif (c’est un signe iconique au sens de C. Peirce), l’*idéogramme* est une combinaison de pictogrammes ayant une signification abstraite, et le *phonogramme* représente un son. La Bibliothèque Nationale de France (BNF) a réalisé un dossier sur les écritures anciennes [1] et Wikipédia propose aussi un portail sur ce thème (<http://fr.wikipedia.org/wiki/Portail:écriture>).

2.2.1.1 Hiéroglyphes égyptiens

À l’origine, les hiéroglyphes égyptiens sont des pictogrammes. Cependant l’écriture a évolué et certains pictogrammes sont devenus des phonogrammes : par exemple le pictogramme “maison” cor-

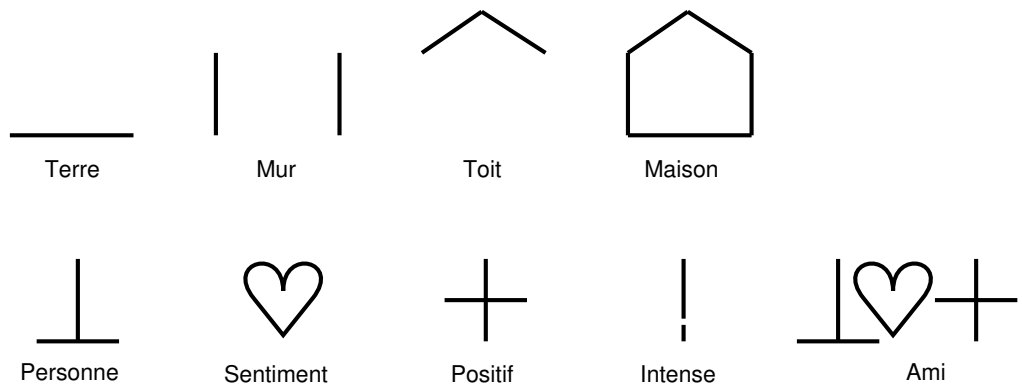


Fig. 2.8 – Quelques symboles de Sémantographie.

En haut : formation d’un symbole dérivé, en bas : symbole composé.

respond au son “PR”. Pour enlever l’ambiguïté, le signe de la bouche est ajouté après lorsque le signe “maison” est utilisé comme phonogramme (figure 2.7). Puis des lettres sont apparues, issues du grec.

La disposition des signes (de gauche à droite, ou de bas en haut,...) est dictée par l’esthétique. Le but de l’écriture est avant tout artistique et religieux : les personnages et événements décrits accèdent à l’éternité, d’où l’importance de la ressemblance avec le sujet représenté. La simplification et la réduction du nombre de signes utilisés pour l’écriture n’est pas recherchée : les scribes sont peu nombreux et souhaitent garder leur pouvoir.

2.2.1.2 Sinogrammes

Les sinogrammes sont la plus ancienne écriture à être encore utilisée aujourd’hui. Elle comprend 55000 caractères, dont 3000 sont d’usage courant. Les caractères occupent la forme d’un carré et appartiennent à différentes catégories :

Les figures simples pictogrammes ou symboles,

Les figures composées les différents éléments voient leur taille réduite pour tenir à l’intérieur du carré :

Les agrégats logiques le signe du soleil, doublé, signifie “brillant” ; placé sous ou sur le signe de l’arbre, il signifie respectivement l’obscurité ou la lumière.

Les complexes phoniques 2 signes sont associés, l’un indique le sens et l’autre la prononciation : par exemple le signe de l’arbre associé à différents signes correspond aux objets en bois (rateau, table, planche,...). Ce sont les plus fréquents.

Cette écriture se caractérise aussi par l’absence d’article, de pluriel, de ponctuation (dans certaines versions) et de conjugaison des verbes. Elle est utilisée pour écrire le chinois, mais aussi utilisée en partie pour écrire le coréen, le vietnamien et le japonais, la lecture dans différentes langues étant possible grâce à la nature non phonétique des figures simples. Cela montre qu’il est possible de réaliser une écriture indépendante de la langue parlée, à condition d’exclure tout élément phonétique.

Si en théorie ces écritures peuvent être comprises sans apprentissage, en pratique le nombre très important de pictogrammes (>1000) et la tendance à les styliser en symboles font que ce n’est pas le cas. De plus, des pictogrammes très différents sont utilisés pour représenter la même chose dans différentes écritures (par exemple l’eau est représentée chez les Égyptiens par une vague en mouvement, chez les Chinois par une évocation du courant et chez les Aztèques par la couleur bleue à l’intérieur d’un récipient).

2.2.1.3 Langages graphiques “universels”

Beaucoup plus récemment, plusieurs tentatives ont été faites pour réaliser un **langage graphique “universel” à base de pictogrammes**. Les Isotypes (*International System Of Typographic Picture*

Education) de O. Neurath (<http://www01.heise.de/tp/english/inhalt/co/2173/2.html>) sont un échec en tant que langage, mais ils sont à l'origine des pictogrammes utilisés dans les aéroports ou les gares. La Sémantographie créée par C.K. Bliss (voir figure 2.8) souffre des mêmes problèmes que les hiéroglyphes ou les caractères chinois. La mémoire de l'éléphant [76] est une approche artistique, principalement utilisée en milieu scolaire.

D'autres tentatives visent à faire communiquer deux personnes de langues différentes *via* Internet (CD-Icon, VIL [91], Miracle [96], Unideo <http://www.unideo.com/>). Ces langages utilisent une interface informatique pour éviter d'avoir à dessiner les pictogrammes à la main : par exemple dans VIL, l'utilisateur choisit un modèle de phrase (par exemple "X donne Y à Z") puis remplit les trous avec les pictogrammes qu'il choisit dans une liste. Bien qu'évitant les problèmes de stylisation des pictogrammes en symboles, ces approches n'ont pas été de réels succès car trop limitées dans l'expression de concepts abstraits.

2.2.2 Pictogrammes standardisés ou recommandés

De nombreux pictogrammes existants ont été standardisés ou "recensés", pour favoriser la réutilisation plutôt que la création de nouveaux pictogrammes.

2.2.2.1 Le jeu de pictogrammes général de l'ISO

La norme ISO 7001 définit le jeu de pictogrammes général présents dans les gares ou les aéroports (toilette, taxi, ascenseur, interdiction de jeter des ordures,... 57 pictogrammes en tout). Dans sa thèse, P. Vaillant [150] réalise une analyse sémiotique de ces pictogrammes et des normes ISO correspondantes.

Il existe aussi **une recommandation ISO pour la création de pictogrammes** (norme IEC 80416-1 : 2001). Cette méthode considère chaque pictogramme comme un "texte non décomposable". Elle suit les étapes suivantes :

1. Réunir un groupe de futurs utilisateurs des pictogrammes (par exemple des médecins dans le cadre qui nous intéresse).
2. Pour chaque concept à représenter, chaque utilisateur doit dessiner une icône.
3. Regrouper les différentes propositions et éliminer les doublons. Les propositions qui représentent les mêmes éléments mais avec des styles de dessin différents seront regroupées ensembles. Par exemple, nous pouvons imaginer les propositions suivantes pour représenter les poumons : un seul poumon, deux poumons, deux poumons avec la trachée entre les deux, ou un symbole comme la lettre 'P'.
4. Demander à chaque utilisateur de classer par ordre de préférence les différentes icônes proposées pour chaque concept.
5. Il est ensuite possible d'évaluer les pictogrammes obtenus sur un autre groupe d'utilisateurs. Cette dernière étape ne figure pas dans la norme ISO, mais a été proposé par De Carolis *et al.* [36] (voir section 2.2.4.3 pour une description des résultats des travaux de De Carolis *et al.*).

2.2.3 Pictogrammes et langages graphiques dans la vie de tous les jours

Plusieurs jeux de pictogrammes ou langages graphiques ont connu un grand succès et sont présents dans notre vie quotidienne.

2.2.3.1 Signalisation routière

La signalisation routière recourt à un langage graphique très simple, utilisé pour signifier des ordres (par exemple limitation de vitesse), des dangers (attention aux virages) ou des informations (directions, commodités) aux conducteurs (voir figure 2.9). Ce langage obéit à deux objectifs : les panneaux doivent être reconnus rapidement, tout en étant indépendants de la langue (pour les touristes



Fig. 2.9 – Exemples de panneaux de signalisation routière (d’après <http://www.securiteroutiere.gouv.fr>).

De gauche à droite : virage à gauche, interdiction aux véhicules agricoles et présence de travaux.

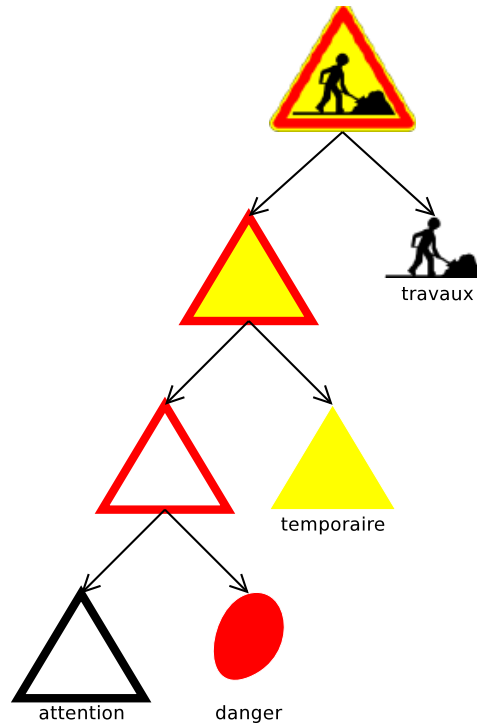


Fig. 2.10 – Structure sémiotique d’un panneau de danger d’après [97].

étrangers). Il fait l’objet d’un apprentissage lors du code de la route, au cours duquel les symboles arbitraires du langage doivent être mémorisés.

Plusieurs auteurs ont tenté d’analyser la structure des panneaux routiers et de définir un **formalisme** pour représenter cette structure ; nous retiendrons celui de J.-G. Meunier [97] qui propose un formalisme graphique inspiré des grammaires génératives de Chomsky. Il décompose le signe en *caractères*, chacun étant une variable rétinienne du signe ou d’une partie de celui-ci. À chaque caractère correspond un *rôle*, c’est-à-dire une signification. Un même caractère peut jouer des rôles différents ; dans ce cas c’est la position des éléments qui permet de faire la distinction (par exemple la distinction entre couleur de fond et couleur d’avant plan). Enfin, des *opérateurs graphiques* permettent de combiner les caractères entre eux : par exemple l’opérateur “colorier” prend deux opérands, une forme et une couleur, et applique la couleur à la forme.

Pour les panneaux routiers, 3 caractères sont utilisés : la couleur, la forme et le pictogramme (voir exemple figure 2.10) :

un pictogramme

une forme géométrique colorée qui entoure le pictogramme et indique le type :

un cercle ordre, prescription

un cercle bleu obligation

un cercle rouge parfois barré interdiction



Fig. 2.11 – Pictogrammes de produits chimiques dangereux (d’après <http://www.aimt67.org/symboles/etiquetage.htm>).

De gauche à droite : produit inflammable, comburant, explosif, corrosif, dangereux pour l’environnement, irritant / nocif, toxique.

un triangle sur la base rouge danger, risque

un carré information

un fond

un fond jaune panneau temporaire

un fond blanc panneau permanent

Quelques panneaux sont des exceptions à la règle (sens interdit, stop, céder le passage ou route prioritaire par exemple; ce sont des signes purement symboliques dont le sens ne peut pas être deviné *a priori*). Cela concerne en particulier les panneaux qu’il est intéressant de pouvoir reconnaître uniquement à leur forme, pour les automobilistes arrivant en sens contraire et voyant le panneau de dos, par exemple le stop.

2.2.3.2 Étiquetage des produits chimiques

Une part importante des connaissances sur le médicament concerne **le danger** (contre-indications, interactions médicamenteuses, certains effets indésirables). La représentation du danger a été beaucoup étudiée pour mettre en garde l’utilisateur contre les risques possibles en cas d’ingestion, d’inflammation,... de produits chimiques (exemple figure 2.11) [133]. Dans le cas de produits d’entretien domestique, les mises en garde doivent être comprises par tous, indépendamment de la langue et sans apprentissage préalable (contrairement à la signalisation routière); les signes arbitraires sont donc à proscrire.

La perception du danger dépend de nombreux facteurs liés à l’utilisateur comme la familiarité avec le produit, l’âge et le sexe (les femmes lisent et suivent plus souvent les avertissements). Si l’information est perçue comme familière, déjà connue de l’utilisateur, elle risque d’être négligée. Au contraire, si elle apparaît absurde à l’utilisateur, celui-ci n’en tiendra pas compte. L’idée que l’utilisateur se fait de la gravité des conséquences est importante, ainsi que l’existence d’une alternative plus sûre en cas de danger ou d’interdiction.

La façon dont le danger est représenté joue aussi un rôle important. Une bonne mise en garde doit combiner plusieurs média : la présence de texte sous forme de phrases standard facilite la reconnaissance et la compréhension des mises en garde, et celle de pictogrammes augmente l’observance et la vitesse de réaction. La forme du pictogramme est plus importante que sa couleur (parce que tout ce qui est rouge n’est pas forcément un danger sur un emballage publicitaire?). Les formes conseillées sont le triangle sur la base pour “attention danger”, le cercle pour “obligation”, le cercle barré pour “interdiction” et le carré ou l’ellipse pour “information”, à rapprocher des formes des panneaux routiers. L’échelle de couleur est la suivante (de la couleur la plus dangereuse à la plus inoffensive) : rouge > orange > jaune > bleu > vert > blanc.

2.2.3.3 Icônes de Logiciels et d’Environnements Informatiques Standardisés (ILEIS)

Les ILEIS sont les icônes utilisées par les logiciels informatiques [112]. Il faut faire attention à ne pas confondre “icône informatique” et “signe iconique” : certaines icônes ne sont pas des signes

iconiques mais des index ou des symboles ! C'est par exemple le cas d'une croix rouge ou d'une flèche vers la gauche pour signifier "annuler".

Les icônes *propres* sont spécifiques à un logiciel (par exemple un logo ou l'icône d'une fonction spécifique à ce logiciel), par opposition aux icônes *génériques* qui se retrouvent quasi à l'identique dans beaucoup de logiciels (par exemple les icônes de sauvegarde ou d'impression). Les icônes génériques forment un vocabulaire visuel et participent à une culture informatique commune à de nombreux utilisateurs. Il existe plusieurs catégories d'icônes :

Icônes de marquage elles font partie des outils disponibles pour la mise en texte et l'ergonomie textuelle, et servent à représenter ou accompagner un autre objet ou à illustrer certaines propriétés d'un objet (nouveau, format d'un fichier,...). Ce sont des signes iconiques [8], ou des symboles lorsque cela est impossible (pour les abstractions non métaphoriques) : une page de texte pour représenter un document texte (signe iconique), un engrenage pour représenter un fichier exécutable (symbole).

Icônes de fonction elles ont pour but de représenter une action offerte à l'utilisateur, et sont donc généralement cliquables. Ce sont généralement des index ou des symboles. On distingue :

une commande qui déclenche une action peu ou pas interactive : une disquette pour "enregistrer" ou une imprimante pour "imprimer" (index),...

un outil interactif mis à disposition de l'utilisateur : un carré pointillé pour "sélectionner" (symbole), un pot de peinture pour "colorier" (index),...

Les constituants d'une icône sont des *iconèmes*. Plusieurs modes d'organisation des iconèmes sont possibles :

Simple un seul iconème,

Superposition

Mimétique la superposition mime une activité telle que l'effectuerait une personne, par exemple une loupe qui parcourt un répertoire lors d'une recherche,

Mixte

Conventionnelle la superposition est purement conventionnelle ; l'ordre et la disposition des iconèmes sont arbitraires,

Apposition les iconèmes sont juxtaposés sans que leur disposition ait un sens (icône de correction d'orthographe avec "ABC" et une coche),

Articulation un articulateur (flèche,...) est introduit entre les iconèmes,

Inclusion un iconème est placé à l'intérieur d'un autre,

Condensation un iconème agit sur un autre de façon métaphorique (par exemple en le compressant pour l'icône d'un logiciel de compression).

Il existe plusieurs guides et recommandations pour la réalisation et l'évaluation des icônes [75, 69, 113, 24]. Voici un extrait du guide de conception des interfaces homme-machine du projet GNOME ([10] chapitre 9) :

- les icônes doivent avoir des silhouettes différentes, pour être distinguées les unes des autres par leur seule forme,
- les icônes analogiques (signes iconiques) sont préférées, avec une analogie qui porte sur l'action associée à l'icône (et pas sur le nom du programme par exemple) ; les icônes symboliques sont à la fois arbitraires et sujettes à des variations culturelles même en cas de convention,
- les icônes ne doivent pas comporter de texte ni se baser sur un jeu de mot, en général intraduisible et qui ne permet pas une association aussi rapide entre l'icône et sa signification (par exemple ne pas représenter un navigateur internet par une araignée sur une toile). Dans ces deux cas, il est nécessaire de passer par un niveau de compréhension linguistique pour déduire la signification de l'icône,

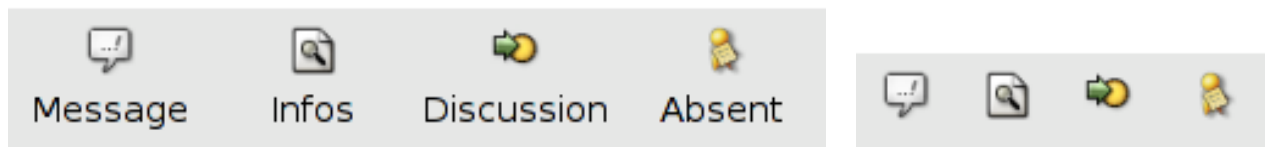


Fig. 2.12 – Icônes + textes ou icônes seules dans le logiciel GAIM.

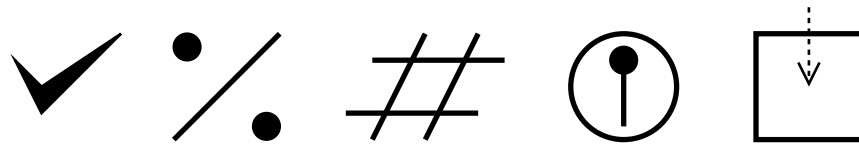


Fig. 2.13 – Exemples de signes médicaux extraits du “Symbol sourcebook” [42].

De gauche à droite : urine, défécation, fracture, coagulation, dispenser.

- les icônes doivent éviter :
 - les images violentes (bombes, explosifs,...) qui intimident l'utilisateur,
 - les parties du corps humain (main, gestes,...) qui peuvent avoir des sens différents selon les cultures,
 - les logos qui n'ont pas de signification pour quelqu'un qui ne connaît pas le programme,
 - les éléments inutiles qui ajoutent du bruit visuel.

Enfin, il faut toujours associer un texte explicatif à une icône [81], car le sens d'une icône est souvent difficile à deviner pour quelqu'un qui ne la connaît pas. Cependant le texte ne doit pas faire partie de l'icône ou se substituer à elle : soit le texte accompagne l'icône, soit il est affiché à la demande (*popup label*). Dans le premier cas, les utilisateurs avancés pourront choisir de masquer les textes (voir figure 2.12).

2.2.4 Pictogrammes et langages graphiques dans le domaine médical

Notre étude bibliographique a mis en évidence un petit nombre de jeux de pictogrammes et de langages graphiques dans le domaine médical. Nous avons d'une part des approches consistant à représenter des données médicales en faisant correspondre différentes variables à des attributs d'une icône (par exemple la largeur et la hauteur d'un rectangle peuvent correspondre au rythme respiratoire et au volume d'air inspiré) [32, 51, 52, 157, 156]. D'autres approches se sont intéressées à des sous-ensembles relativement restreints des connaissances médicales; nous allons en présenter plusieurs de manière plus détaillée.

2.2.4.1 Le Symbol sourcebook

Dans son livre “Symbol Sourcebook”, H. Dreyfuss [42] dresse une liste impressionnante des signes utilisés dans divers domaines (ce que Dreyfuss appelle “symboles” sont des signes selon la classification de Peirce). On constate que ces signes servent principalement à donner des instructions, des ordres ou à mettre en garde, et que d'un domaine à l'autre, les symboles arbitraires sont souvent les mêmes avec des sens totalement différents. Par exemple le triangle sur la base signifie “danger” en général, “courant tri-phasé” en électricité, “chauffer” en chimie, “feu” en alchimie, “extraire” en programmation, “tête” en biologie ou “comportement” en architecture!

Dans le domaine médical, la plupart des signes présentés ne portent pas sur le médicament ni même les connaissances médicales en général : il s'agit de signes indiquant soit le fonctionnement d'appareils de mesure, soit les différents services hospitaliers. Quelques signes correspondent à des concepts médicaux intéressants (voir figure 2.13), cependant ceux-ci sont arbitraires et ne semblent pas d'usage courant en France à notre époque.



Fig. 2.14 – Exemple de pictogrammes de l’AFSSAPS.

Amphotericin B							
Stabilité des solutions Stabilità in soluzione		Stability in solutions Estabilidad en solución		Stabilität der Lösung			
PVC	◆	0,05 & 5 mg/ml	23°C-25°C	☀	24	⊙	569 ?
PVC	◆	0,1 mg/ml	15°C-25°C	☀	24	⊙	876 ?
PVC	◆	0,1 & 0,25 mg/ml	4°C	☀	35	◐	508 ?
PVC	◆	0,2 & 0,5 mg/ml	25°C	☀	120	⊙	435 ?
PVC	◆	0,2 & 0,5 & 1 mg/ml	4°C	☀	120	⊙	435 ?

Fig. 2.15 – Copie d’écran de Stabilis 2, d’après [155].

2.2.4.2 Pictogrammes de l’AFSSAPS

L’AFSSAPS a mis au point quelques pictogrammes destinés à l’étiquetage des boîtes de médicament, à destination du patient [4]. Cependant, ces pictogrammes sont en nombre très limité, et ne respectent pas toujours les conventions. Par exemple, le pictogramme de la figure 2.14 s’inspire des panneaux routiers et utilise une forme triangulaire ; cependant, selon la grammaire des panneaux routiers, ce pictogramme signifie “risque de conduite automobile” et non pas “danger en cas de conduite automobile” !

2.2.4.3 Icônes du projet OPADE

Dans le domaine médical, dans le cadre du projet OPADE, De Carolis *et al.* [36] ont réalisé une série d’icônes destinées à être utilisées dans les barres d’outils de logiciels médicaux. Elles représentent des actions comme “afficher l’historique du patient”, “afficher la monographie du médicament” ou “il y a une contre-indication”. Cependant ces icônes ne permettent pas d’indiquer de quel type de contre-indication il s’agit (par exemple contre-indication avec l’asthme). Il s’agit plus d’un ensemble d’icônes prédéfinies que d’un “langage iconique”, contrairement à ce qui est dit dans l’article.

2.2.4.4 Stabilis 2

Stabilis 2 [155] est une base de connaissance sur la stabilité et la compatibilité des médicaments injectables. Ce logiciel utilise un langage à base d’icônes (environ 150) pour indiquer les classes thérapeutiques des médicaments, les conditions de conservations, les stabilités dans différentes solutions, les incompatibilités,... L’utilisation de pictogrammes permet notamment de s’affranchir de la langue, alors que les sources de connaissance (littérature médicale et livres de référence spécialisés) sont quasi-exclusivement en anglais.

L’initiative est intéressante, cependant Stabilis peine à établir une véritable grammaire graphique rigoureuse : les différents éléments du langage n’ont pas toujours la même signification d’un pictogramme à l’autre. Ainsi, par exemple (voir figure 2.16), l’icône avec un carré rouge barré et un soleil signifie “conserver à l’abri de la lumière”, mais l’icône avec le même carré rouge barré et un



Fig. 2.16 – Exemples d’icônes issues de Stabilis 2, d’après [155].

A gauche : “conserver à l’abri de la lumière”, à droite : “anti-fongique”.

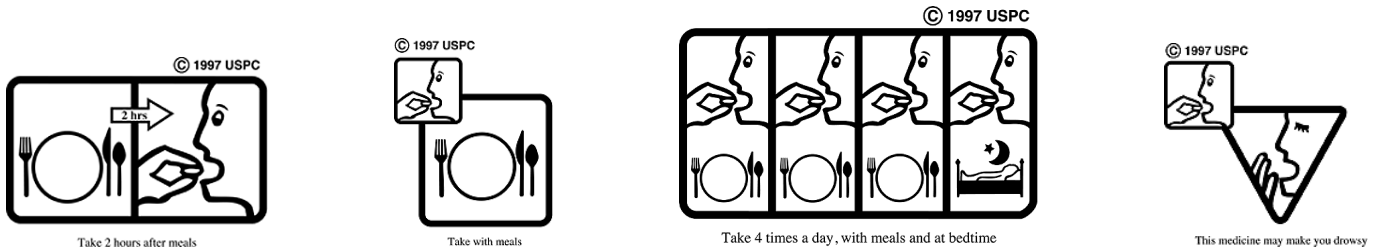


Fig. 2.17 – Exemples de pictogrammes de l’USP, d’après <http://www.usp.org>.

De gauche à droite : prendre deux heures après les repas par voie orale ; prendre pendant les repas par voie orale ; prendre 4 fois par jour, aux repas et au coucher ; risque d’altération de la conscience.

champignon signifie “anti-fongique” et non “conserver à l’abri des champignons”!²

2.2.4.5 Pictogrammes de l’USP

L’USP (*United States Pharmacopeial Convention*) a mis au point une série de 81 pictogrammes pour représenter des instructions, précautions ou mises en garde sur la prise de médicaments à destination du patient (voir figure 2.17 ; <http://www.usp.org/drugInformation/pictograms/>). Ces pictogrammes concernent plus précisément le plan de prise, les effets secondaires principaux, la voie et les précautions d’administration (par exemple lavage des mains, dilution,...), les précautions pour la conservation du médicament, les interactions avec l’alimentation (produit laitier, alcool,...), les contre-indications (grossesse, allaitement,...),... D’autres sont trop généraux pour être vraiment utiles (“lisez la notice”, “ne donnez pas vos médicaments à quelqu’un d’autre”,...). Plusieurs évaluations ont été réalisées, avec des résultats mitigés [149, 65, 141].

Les conventions utilisées sont les suivantes : le pictogramme est entouré d’un cercle barré par un X pour une interdiction, un triangle sur la pointe pour les précautions et un carré ou un rectangle pour une information générale. Un petit carré peut être rajouté en haut à gauche pour indiquer quel aspect est concerné (prise ou conservation du médicament). Chaque pictogramme est accompagné d’une phrase (en anglais).

D’autres jeux de pictogrammes, proches de ceux de l’USP et toujours destinés au patient, ont été recensés dans la littérature [37, 73].

2.2.4.6 UVAL-MED

UVAL-MED [116, 117, 118] est un langage graphique permettant de décrire à l’aide de schémas les principales caractéristiques des pathologies : symptômes,... afin de faciliter le diagnostic. Il comprend un ensemble de pictogrammes de base ainsi qu’une grammaire graphique simple permettant de combiner les pictogrammes. Chaque schéma comprend plusieurs parties, qui montrent une région anatomique (ex. abdomen), un compartiment (ex. sang), ou le patient dans son ensemble (par exemple une personne avec une jupe pour indiquer que la pathologie concerne plus particulièrement la femme).

L’objectif d’UVAL-MED est d’abord pédagogique : il s’agit de faciliter l’apprentissage de la physiopathologie par des étudiants en médecine. UVAL-MED a été évalué sur des étudiants en néphrologie. Les 33 étudiants du groupe de contrôle disposaient de la description textuelle des

²Ces commentaires ont été communiqués aux auteurs de Stabilis 2 qui en tiendront compte dans la prochaine version de leur logiciel.

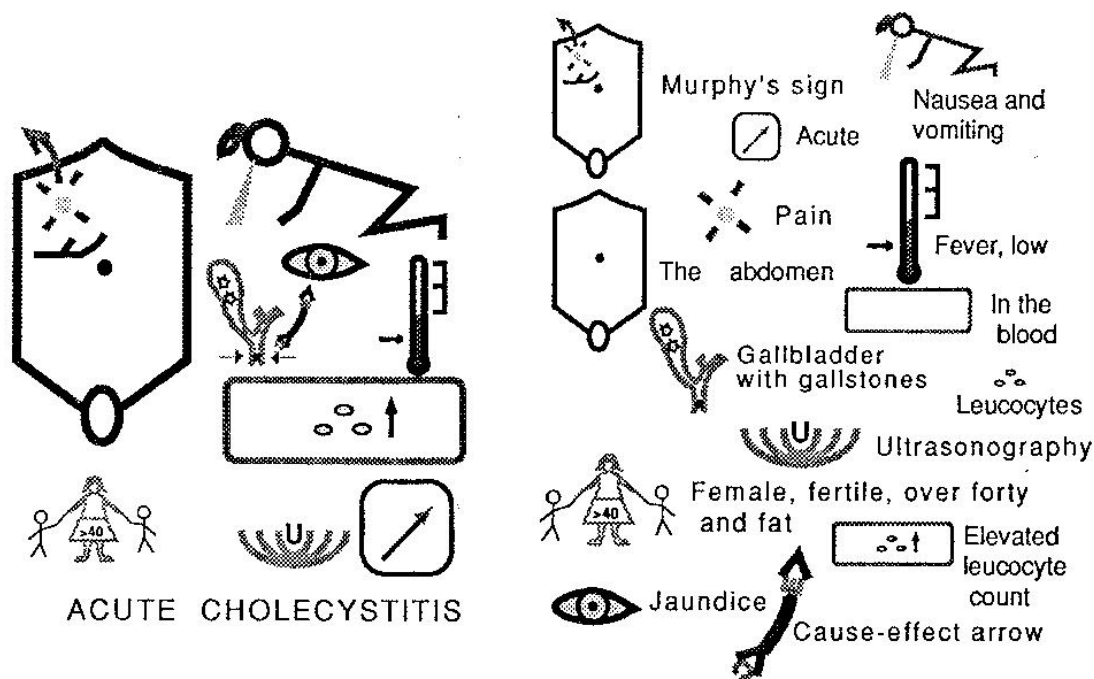


Fig. 2.18 – Représentation d'une pathologie avec UVAL-MED, d'après [117].

pathologies, et les 39 étudiants du groupe de test avaient en plus les schémas UVAL-MED. Les documents étaient retirés après une période d'apprentissage d'une demi-heure, et les étudiants devaient répondre à un questionnaire. Les résultats sont significativement meilleurs lorsque les schémas sont utilisés pour les questions de raisonnement ; la différence n'est pas significative pour les questions de mémoire pure. Par ailleurs, l'acceptation des schémas par les étudiants est très bonne, et ils les jugent plus utiles que les livres ou les autres ressources pédagogiques. Cependant, malgré cela, il semble que ce langage n'ait pas donné lieu à des travaux ultérieurs ni à une industrialisation.

2.2.5 Conclusion sur les langages graphiques

Nous avons vu qu'il existe un grand nombre de langages graphiques, conçus pour des besoins variés, limités à un domaine ou au contraire généralistes. En revanche, **il n'existe pas dans la littérature de méthode pour concevoir ces langages graphiques** ; les seules méthodes existantes concernent la création de pictogrammes, comme les recommandations ISO. Cependant celles-ci considèrent chaque pictogramme comme un "texte non décomposable". Il n'est donc pas possible de ré-utiliser les pictogrammes existants pour en créer d'autres sans qu'il s'agisse de nouveaux pictogrammes, qui doivent faire l'objet d'une évaluation complète. En conséquent, il n'y a pas de grammaire : il s'agit d'un vocabulaire (catalogue de mots) mais pas d'un véritable langage.

Dans le domaine médical, plusieurs jeux de pictogrammes ont été proposés (Stabilis 2, pictogrammes de l'USP,...), mais seul UVAL-MED est un véritable langage proposant une grammaire graphique. Cependant, ce langage n'a pas été poursuivi, et aucun jeu de pictogramme ou langage graphique n'a été proposé pour accélérer l'accès aux connaissances médicales par les professionnels de santé.

Les formalismes utilisés pour représenter les grammaires des langages, comme les BNF, *Backus-Naur Form*, ou les diagrammes de Conway, sont insuffisant pour représenter les grammaires des langages graphiques, car ces formalismes s'appliquent à des langages produisant des phrases qui sont des suites de symboles, alors que nous avons vu qu'un langage graphique peut combiner des éléments graphiques de différentes manière, et pas seulement les uns à la suite des autres. De plus, le champ des grammaires formelles se limite à l'expression de la syntaxe, alors que lors de la définition d'un langage graphique, c'est d'abord la sémantique de la grammaire qui nous intéresse.

Dans la littérature, plusieurs formalismes ont été proposés pour représenter la grammaire d'un

langage graphique sans que l'un d'eux semble s'imposer comme standard de fait ; nous avons présenté celui de J.-G. Meunier à la section 2.2.3.1. Ces formalismes sont soit de nature textuelle, ce qui les rend peu adaptés à la représentation de langage graphique, soit de nature graphique et basé sur la décomposition des éléments du langage graphique (c'est le cas de celui de J.-G. Meunier) ; cependant ces formalismes sont assez limités : par exemple il n'est pas possible d'indiquer le sens associé à l'absence d'un élément, car on ne peut pas représenter l'absence de cet élément.

2.3 La visualisation d'information

La *visualisation d'information* est l'étude de la représentation graphique d'informations quelles qu'elles soient, et notamment des informations abstraites pour lesquelles il n'y a pas de représentation graphique "évidente", c'est-à-dire qui n'ont pas de propriétés spatiales/géométriques (contrairement à la visualisation scientifique, par exemple les cartes de température pour la météo). La plupart des connaissances sur le médicament font partie de cette catégorie (à l'exception des informations purement anatomiques, pour lesquelles il y a une localisation spatiale sur le corps humain).

L'objectif est double :

- rendre l'information plus accessible (en cas de grande quantité de données) et permettre l'exploration à différents niveaux d'abstraction,
- permettre l'analyse visuelle des données, en exploitant les capacités de la vision (segmentation, reconnaissance de motifs ou de formes,...), on parle alors de "*data-mining* visuel" [30].

La visualisation d'information est à la fois concurrente [90] (second objectif) et complémentaire [101] (premier objectif) des techniques de fouille de données automatique (*data mining, knowledge discovery,...*). La visualisation d'information implique l'utilisateur de manière active dans le raisonnement, ce qui favorise l'acceptation de ces techniques par l'utilisateur [90, 30].

Les différentes techniques de visualisation d'information sont généralement classées selon la structure des informations visualisées (texte, graphique, matrice objet-attribut, arbre, réseau, indices de similarité) [28, 23, 66, 140] ; dans le domaine médical, une classification de ces techniques a été proposée en fonction du mode de visualisation [144].

L. Chittaro [29] a réalisé une revue des applications à base d'approches graphiques dans le domaine médical, et G. Kwok-Chu Ng [82] une revue des approches graphiques destinées à la représentation d'ontologie. En ce qui concerne les visualisations tridimensionnelles, C. Bruley [18] a réalisé une revue des approches existantes et J. Moline [100] une revue des approches à base de réalité virtuelle en médecine. Pour ma part, j'ai réalisé lors de ma 5ème année recherche de pharmacie un visualisateur 3D pour la télésurveillance de personnes âgées, afin d'éviter l'utilisation de caméras jugées trop intrusives [129, 84]. Enfin, K. Andrews [6] a proposé une liste quasi-exhaustive des systèmes de visualisation d'information existants.

Nous allons tout d'abord présenter le *Fisheye*, un procédé utilisé par de nombreuses techniques de visualisation pour ajouter de l'interactivité, puis nous étudierons les différentes techniques de visualisation d'information, en suivant la classification habituelle, en fonction de la structure des informations visualisées.

2.3.1 *Fisheye*

Le *Fisheye* (oeil de poisson, [60, 66]) sépare les informations à visualiser en deux catégories selon le point de vue de l'utilisateur : le *focus* et le *contexte*, et affiche plus d'information dans le focus que dans le contexte. La limite entre le focus et le contexte peut être franche, ou bien progressive. Puisque l'utilisateur doit indiquer le focus, le *Fisheye* permet de rendre la visualisation interactive, ce qui la rend plus efficace. Le principe de *Fisheye* est associé à de nombreuses techniques de visualisation que nous verrons par la suite.

Le *Fisheye* peut être défini par une fonction DOI (*Degree Of Interest*, degré d'intérêt, [60]) qui pour chaque élément retourne son degré d'intérêt ; plus celui-ci est élevé et plus l'élément est proche

```

def start_server():
    global SERVER

    server_prog = os.path.join(os.path.dirname(sys.argv[0]), "balazard")
    server = os.spawnl(os.P_NOWAIT, server_prog, server_prog, "--port", globdef.PORT)
    #server = os.popen2(os.path.join(os.path.dirname(sys.argv[0]), "balazard"))
    ...
def stop_server():...
def build_scene(idler, level, character):...
def clean_mem():...
...

```

Fig. 2.19 – Exemple de repliement de code dans GNU Emacs (d’après <http://mail.python.org/pipermail/python-list/2002-May/106656.html>).

À gauche, la fonction Python `start_server` est dépliée, à droite elle a été repliée. Les repliements sont indiqués par des “...”.

du focus. Un exemple de DOI est : $DOI(X) = Importance(X) - DistanceEntre(X, PointDeVue)$, où `PointDeVue` est le point de vue de l’utilisateur. Il existe deux types de *Fisheye* :

***Fisheye* filtrant** La séparation entre focus et contexte a lieu avant le processus de visualisation, et les éléments dont le DOI est trop faible ne sont pas affichés. La version la plus simple est le zoom (qui consiste à cacher le contexte pour laisser plus de place au focus [99, 9]).

***Fisheye* déformant** Plutôt que de filtrer l’affichage, il s’agit de le déformer pour attribuer plus d’espace aux éléments dont le DOI est élevé [132]. La séparation entre focus et contexte a lieu après le processus de visualisation et consiste à sélectionner des portions de l’écran (pixels). La limite entre focus et contexte n’est pas nette ; le passage d’une région à l’autre se fait par une transition progressive.

L’utilisation de la perspective dans une visualisation en 3D permet d’obtenir un effet de *Fisheye* déformant [131] : les objets les plus proches sont affichés en plus grand, les plus éloignés en plus petit, et les objets trop loin ou situés derrière un autre objet sont invisibles. Comme ces propriétés se retrouvent dans le monde réel, elles semblent naturelles à l’utilisateur.

2.3.2 Méthodes de visualisation de textes et de listes

2.3.2.1 Repliement (*folding*)

Cette technique de *Fisheye* filtrant est surtout utilisée par les environnements de programmation (*code folding*, figure 2.19) et permet de réduire le bruit visuel. Il s’agit de n’afficher que les en-têtes des fonctions et méthodes d’un fichier de code, et d’afficher le contenu de la fonction / méthode en fonction leur intérêt (le DOI correspond, selon les systèmes, à la demande de l’utilisateur (clic sur l’en-tête dans Emacs) ou à l’importance du contenu [60]). Le contexte est l’ensemble des en-têtes, et le focus est formé par les corps des fonctions/méthodes dépliées.

2.3.2.2 *Fisheye* déformant

Plusieurs méthodes à base de *Fisheye* déformant ont été proposées pour faciliter la visualisation de textes ou de listes. La lentille de document (*document lens*, figure 2.20) [130] agrandit une partie du texte tout en gardant le contexte visible (contrairement à ce qui se produit avec une loupe, où les informations proches de la partie agrandie sont cachées par la loupe). Le mur fuyant (*perspective wall*, figure 2.21) [95] consiste à plaquer une liste sur un mur composé de trois faces : une face parallèle à l’écran qui affiche un maximum d’information, et deux faces de part et d’autre de la première qui s’éloignent en profondeur, et sur lesquelles les informations sont compressées à cause de la perspective, ce qui permet d’en afficher plus. Tous deux ont été développés par les laboratoires Xerox afin de représenter plus de détails dans le focus au centre de l’écran et moins dans le contexte.

2.3.2.3 *Greeking*

Le *greeking* est une technique pour représenter du texte de manière très concise : chaque ligne de texte est remplacée par une ligne horizontale de longueur proportionnelle au nombre de caractères. Il

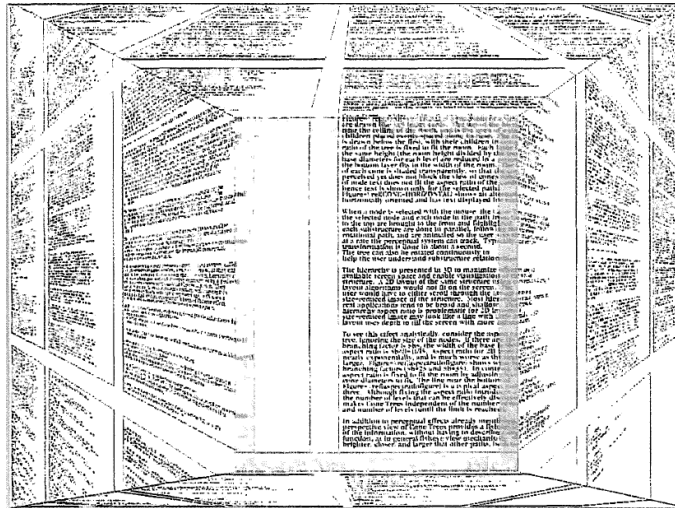


Fig. 2.20 – Lentille de document.

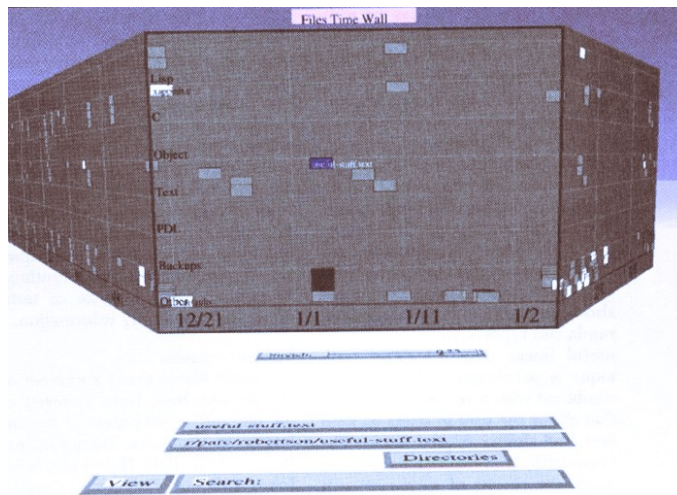


Fig. 2.21 – Mur fuyant représentant des fichiers triés par date.

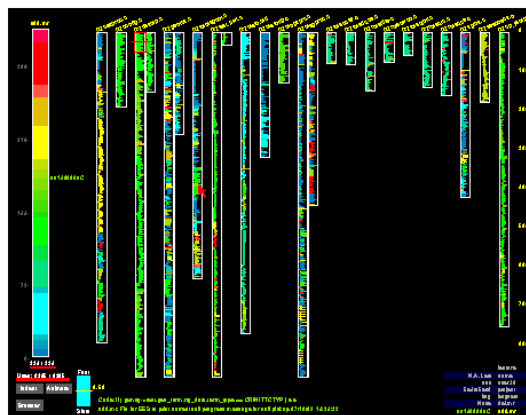


Fig. 2.22 – Visualisation de code avec SeeSoft.

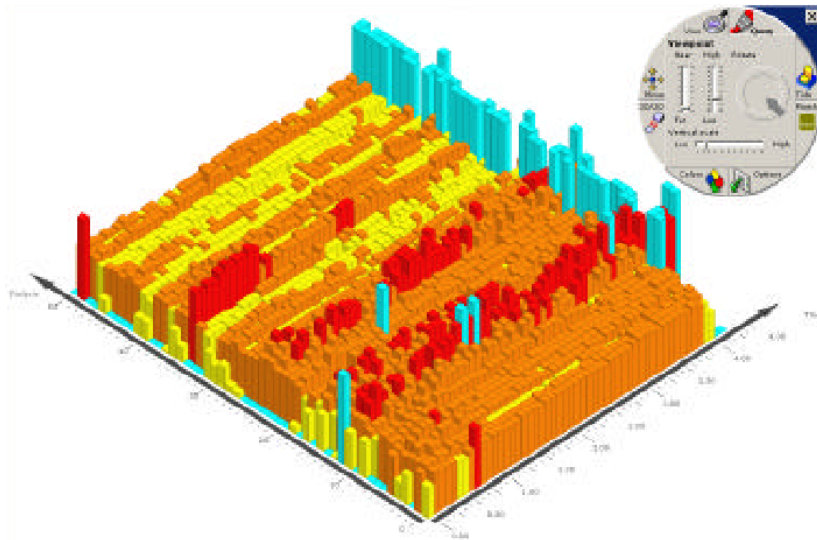


Fig. 2.23 – IPBC (*Interactive Parallel Bar Charts*) [30] représentant des données d'hémodialyse. Sur l'axe X (à droite) : le temps, Z (à gauche) : les différentes séries (1 série=1 paramètre fourni par l'appareil à hémodialyse), Y (en hauteur) : la valeur de la série.

	Escherichia Coli
Acide fusidique	<div style="width: 100%; height: 15px; background-color: red;"></div>
Aminosides	<div style="width: 100%; height: 15px; background-color: green;"></div> <div style="width: 10%; height: 15px; background-color: orange;"></div>
Céphalosporine 1G	<div style="width: 30%; height: 15px; background-color: green;"></div> <div style="width: 20%; height: 15px; background-color: orange;"></div> <div style="width: 50%; height: 15px; background-color: red;"></div>

Fig. 2.24 – Visualisation de spectres antibactériens [44, 43].

n'est plus possible de lire le texte, mais il est toujours possible de reconnaître des motifs, des passages plus ou moins denses ou des titres/ alinéa : seule la silhouette du texte est conservée.

Cette technique est utilisée dans SeeSoft (figure 2.22) [49] pour la représentation de code source informatique. La couleur du texte est déterminée par une caractéristique de la ligne, par exemple la dernière date de modification du passage correspondant.

2.3.3 Méthodes de visualisation de données multidimensionnelles et matrices objet-attribut

2.3.3.1 Graphiques

Les graphiques "traditionnels" (camemberts, bâtons, lignes brisées, nuages de points,...) peuvent être utilisés, en 2D ou en 3D. Par exemple, L. Chittaro *et al.* [30] ont proposé une application à base d'histogrammes (*Parallel Bar Charts*) pour le monitoring et la *data-mining* visuel des données acquises par les appareils d'hémodialyse (figure 2.23). Différents outils permettent de changer l'angle de vue, de modifier les couleurs, de cacher certaines rangées de barres, ou d'ajouter un "niveau d'eau" pour mettre en évidence les barres qui dépassent hors de l'eau. Le niveau d'eau peut être incliné, pour aider à détecter des changements de pente.

C. Duclos [44, 43] a proposé un système de visualisation des spectres anti-bactériens des antibiotiques (figure 2.24). Pour chaque combinaison antibiotique-bactérie, le spectre est représenté par une barre découpée en trois parties de couleur correspondant au niveau de résistance bactérienne (vert : sensible, orange : intermédiaire, rouge : résistante). La longueur de chaque partie est proportionnelle à la prévalence du niveau de résistance.

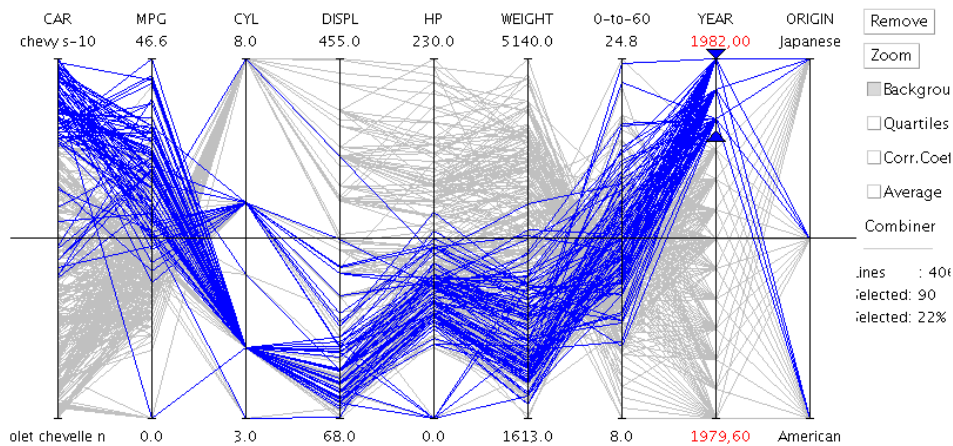


Fig. 2.25 – Exemple de coordonnées parallèles, représentant des modèles de voitures disponible en applet sur <http://www.cs.uta.fi/~hs/pce/>.

La sélection des modèles récents (en bleu, sélection sur l’avant dernier axe) permet de voir une tendance à la baisse du poids du véhicule et du nombre de cylindres.

2.3.3.2 Coordonnées parallèles

Les coordonnées parallèles de A. Inselberg peuvent représenter dans un espace 2D des données ayant un nombre quelconque de dimensions. Il y a un axe vertical par dimension, et chaque élément est représenté par une ligne brisée reliant les différents valeurs qui lui correspondent sur les axes.

Cependant, lorsque plusieurs lignes se rencontrent au même point (par exemple sur l’axe CYL(indre) de la figure 2.25), il y a confusion et il n’est pas possible de faire correspondre les lignes qui arrivent à celles qui repartent. Plusieurs solutions ont été proposées : permettre à l’utilisateur de sélectionner une plage de valeurs sur un axe et changer la couleur des lignes qui passent à travers cette plage, remplacer les lignes par des courbes lissées,... Notons aussi que l’ordre des colonnes peut atténuer ce problème (par exemple les colonnes avec un petit nombre de valeurs discrètes doivent être placées aux extrémités, comme l’axe ORIGIN(e)), et rendre plus ou moins facile l’interprétation des interactions (plus les colonnes sont proches, plus cela est facile). Ce problème se rencontre particulièrement pour les valeurs quantitatives ; par exemple dans la figure précédente il est impossible de répondre à la question “les voitures récentes sont-elles en majorité japonaises?”, même avec la sélection. En effet, il y a chaque année au moins une nouvelle voiture de chacune des trois origines.

2.3.3.3 Glyphes

Un glyphe est un signe graphique dont l’apparence est déterminée par l’objet qu’elle représente. Chaque propriété (ou dimension) de l’objet est associée à une propriété du glyphe (longueur d’une partie du glyphe, couleur, forme,...) de manière arbitraire. Par exemple, InfoBug [31] représente les bogues d’un logiciel sous la forme d’un insecte, dont la longueur de l’aile gauche dépend du nombre de lignes de code concernées, et la longueur de l’aile droite du nombre d’erreurs (figure 2.26) ; d’autres propriétés définissent la longueur des antennes, le nombre de points,... N. Osawa [106] représente les classes d’un langage de programmation objet par des glyphes abstraits.

Une approche particulière est le glyphe en étoile (“*stardinate*”, *star plot*, *sunflower plot*) [89, 55], inspiré des coordonnées parallèles (section 2.3.3.2). Le glyphe est construit comme un graphique de coordonnées parallèles, mais au lieu d’être parallèles les axes se coupent en un point, le centre de l’étoile (figure 2.27).

Le glyphe peut être abstrait (il ne ressemble à rien d’existant) ou bien métaphorique (il ressemble à quelque chose mais pas à l’objet qu’il représente). Les visages de Chernoff (*Chernoff’s faces*) [27]³ sont des glyphes métaphoriques en forme de visage humain ; les différentes propriétés du glyphe étant la taille des yeux, la longueur du nez, la position de la bouche,... Il a été prouvé que la vision

³Voir applet de démonstration : <http://people.cs.uchicago.edu/~wiseman/chernoff/>

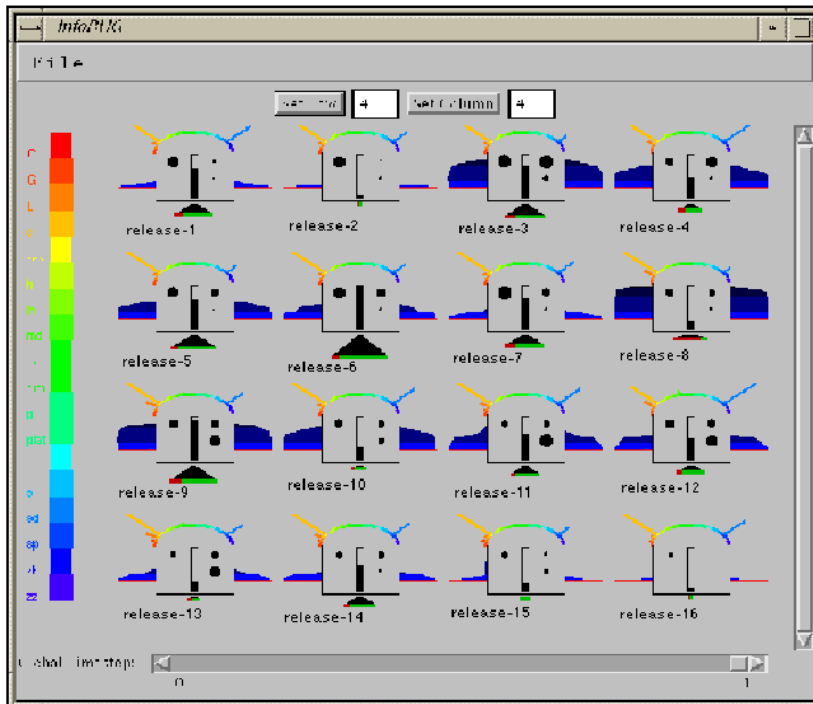


Fig. 2.26 – 16 logiciels représentés par InfoBug.

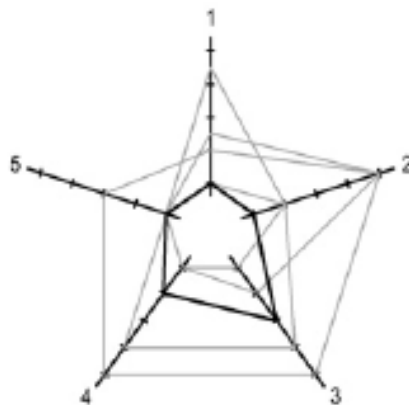


Fig. 2.27 – Glyphe en étoile.

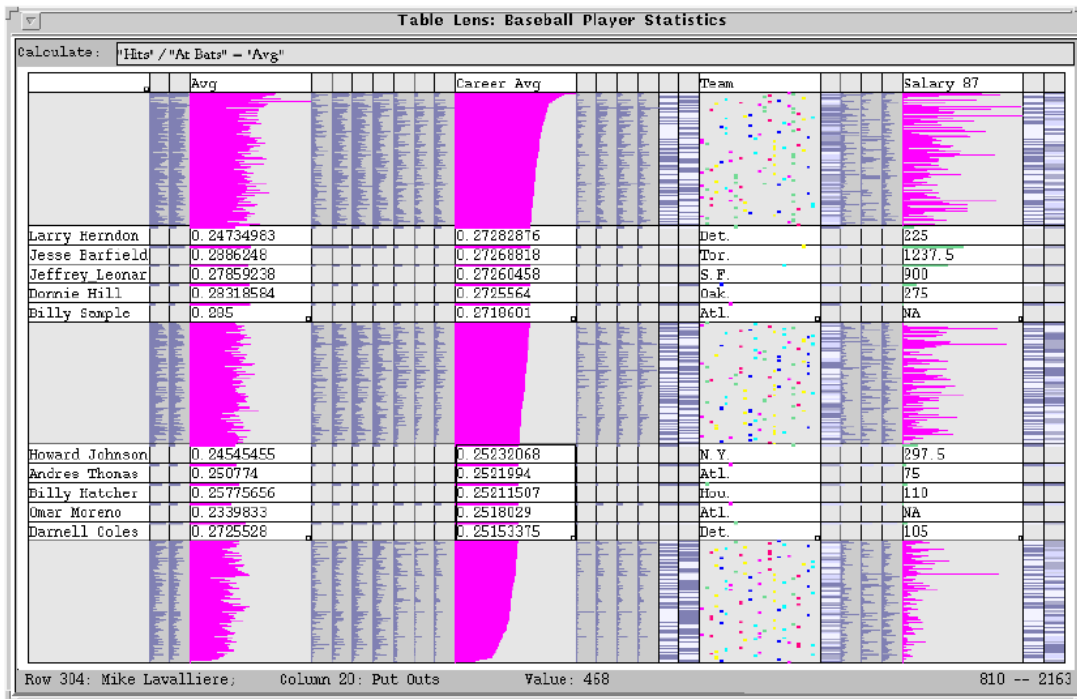


Fig. 2.28 – Lentille de tableau montrant la visualisation d’une base de données de joueurs de base-ball. Ici, nous avons deux zones focales horizontales (de 5 lignes chacune) et cinq verticales (d’une colonne chacune).

humaine est très performante pour percevoir les visages humains, ce dont tirent parti les visages de Chernoff. Cependant, les visages humains expriment aussi des sentiments (joie, peur, tristesse,...). Ces sentiments viennent souvent perturber l’utilisateur ; par exemple si l’on représente des médicaments par des visages de Chernoff, le médecin risque d’être attiré par les visages souriants ou paisibles. Or celles-ci ne correspondent pas forcément aux “meilleurs” médicaments, et les interactions entre les différents composants du visage sont extrêmement complexes.

Le glyphe n’a *a priori* rien à voir avec l’objet qu’il représente. L’interprétation des glyphes est donc difficile, cependant les glyphes sont très pratiques pour la recherche d’objets similaires : en effet, les objets proches ont des glyphes qui se ressemblent.

2.3.3.4 Lentille de tableau

Un tableur permet d’afficher un maximum d’environ 660 cellules à l’écran ; une lentille de tableau (*table lens*) permet de multiplier ce nombre par 30 et jusqu’à 100 en utilisant des techniques de *Fisheye* [114, 125] (figure 2.28). Le principe de base est d’afficher plus d’information dans les cellules du focus : par exemple les cellules du contexte sont réduites à un trait horizontal d’un pixel de hauteur dont la longueur est proportionnelle à la valeur de la cellule, les cellules du focus sont agrandies et contiennent la valeur chiffrée. Les colonnes ont un fonctionnement identique, et sont agrandies lorsqu’elles sont sélectionnées. InfoZoom offre des possibilités similaires [142], et de nombreux systèmes de visualisation s’inspirent des tableurs [143, 103].

Dans le domaine médical, Dopamine (*Drug Ontology Production And MaIntenance eNvironE-ment*, [161], figure 2.29) est un outil pour visualiser les propriétés des médicaments inspiré des *tables lens*. Les différents médicaments correspondent à l’axe horizontal, et les propriétés à l’axe vertical. Les propriétés sont affichées en liste arborescente, ce qui permet de regrouper les propriétés similaires pour gagner en abstraction ou au contraire de les séparer. Dopamine repose sur la *Drug Ontology* qui utilise la classification de Galen, et il est orienté vers la conception et la maintenance de l’ontologie, et plus généralement l’acquisition des connaissances, mais pas vers une utilisation clinique.

Features:		6 dissections				
		Clofibrate BEZAFIBR PROFIBR CLOFIBR FENOFIBR GEMFIBR				
- contraindication						
HAS_FEATURE	avoid					
WHICH_IS	agitation state in elderly					
WHICH_IS	alcoholism					1
WHICH_IS	basal ganglia disease					
WHICH_IS	breast-feeding	1	1	1	1	1
WHICH_IS	gall bladder disease	1	1	1	1	
WHICH_IS	gallstones					1
+ WHICH_IS	hepatic impairment	1	1	1	1	1
WHICH_IS	hypoalbuminaemia	1	1	1	1	
WHICH_IS	nephrotic syndrome	1	1	1	1	
WHICH_IS	pregnancy	1	1	1	1	1
WHICH_IS	primary biliary cirrhosis	1	1	1	1	
+ WHICH_IS	renal impairment	1	1	1	1	

Fig. 2.29 – Copie d’écran de DOPAMINE.

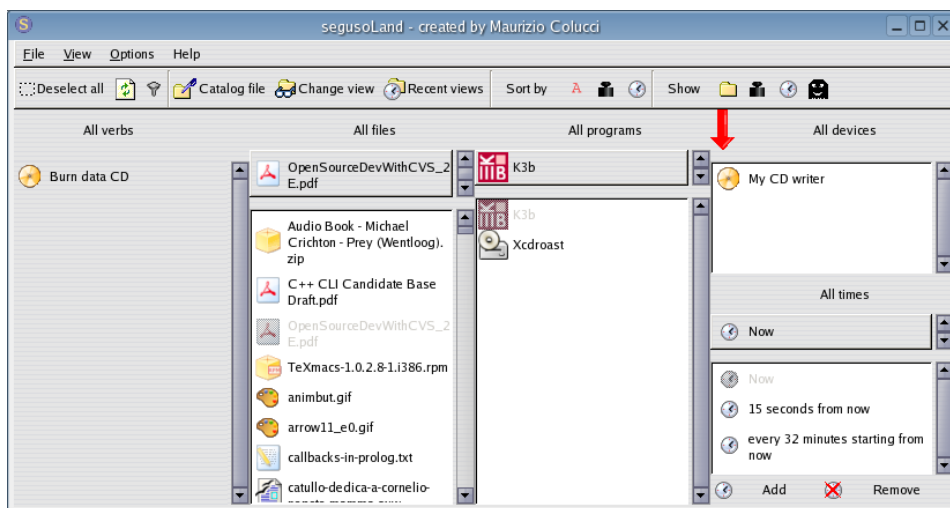


Fig. 2.30 – Copie d’écran de SegusoLand.

2.3.3.5 Réduction réciproque de liste

SegusoLand (figure 2.30, <http://segusoland.sourceforge.net/>) est une interface utilisateur pour rendre l’informatique accessible à tous en combinant dans une seule interface les fonctions de menu de démarrage, de gestionnaire de fichier et d’assistant à des tâches fréquentes. L’interface se compose de 5 listes comprenant les verbes possibles (“graver un CD”, “voir des images”, “écouter de la musique”,...), les fichiers, les logiciels, les périphériques et les dates. L’utilisateur sélectionne dans chaque liste les éléments qui correspondent à l’action souhaitée, par exemple le verbe “graver un CD”, les fichiers à graver, un logiciel de gravure et le périphérique “graveur de CD”, dans n’importe quel ordre. A chaque fois qu’un élément est sélectionné, tous les éléments qui n’ont pas de sens avec sont cachés, par exemple sélectionner un logiciel de gravure cachera le verbe “voir des images” et le périphérique “imprimante”.

Ce principe de *Fisheye* filtrant est appelé réduction réciproque de liste (*reciprocal list narrowing*). Les objets visualisés sont les combinaisons verbe-fichier-programme-périphérique-date pertinentes, ce qui correspond à une matrice objet-attribut avec 5 attributs (un par colonne). Il faut pouvoir lister l’ensemble des valeurs possibles des attributs, ce qui exclut les valeurs quantitatives.

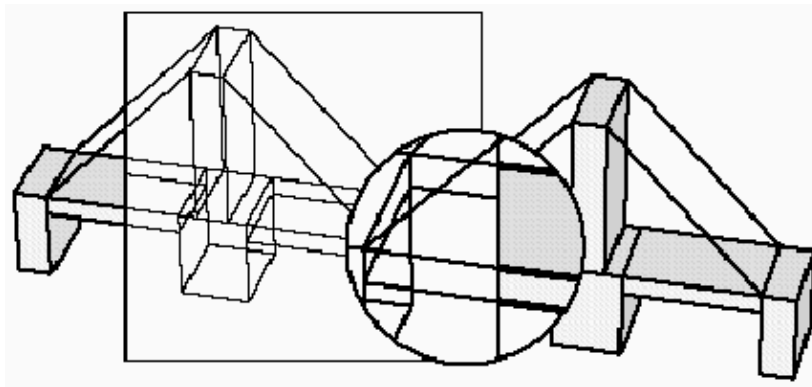


Fig. 2.31 – Lentilles magiques dans un logiciel de dessin.

La lentille ronde est une loupe, la lentille rectangulaire affiche en fil de fer.

2.3.3.6 Requêtes

Il existe plusieurs méthodes graphiques pour faciliter les requêtes au sein de matrices objet-attribut (voir [68] pour une revue). Les requêtes dynamiques de C. Ahlberg [5] permettent à l'utilisateur de spécifier sa requête en utilisant une interface graphique, et les résultats sont mis à jour en temps réel. Les requêtes à facettes [68] permettent d'effectuer plusieurs requêtes simples et de leur associer un nom ; l'utilisateur peut ensuite combiner ces requêtes entre elles, ou effectuer de nouvelles recherches au sein des précédentes.

Avec les lentilles magiques [56], l'utilisateur sélectionne une région à l'écran puis effectue une opération sur cette région : filtrage d'éléments, déformation de l'affichage,... Cette technique utilisée à l'origine par des logiciels de dessin et qui proposait des loupes, des vues transparentes,... (figure 2.31) a été ensuite utilisée pour effectuer graphiquement des requêtes. Les lentilles magiques peuvent potentiellement être utilisées avec n'importe quel type de données, et pas seulement des matrices objet-attribut, puisqu'elles s'appliquent *après* visualisation.

G. Kwok-Chu Ng [82] propose un outil de visualisation de l'ontologie GRAIL à base de lentilles magiques, cependant il conclut que les lentilles magiques sont peu adaptées à la visualisation d'ontologies. Leur utilisation doit être limitée aux applications où il est fréquent de restreindre une requête à une région de l'espace.

2.3.4 Méthodes de visualisation d'arbres

I. Herman *et al* [71] ont réalisé une revue complète des méthodes proposées pour visualiser des arbres ou des réseaux. Nous présentons ici les principales approches.

2.3.4.1 Diagrammes

Il existe plusieurs algorithmes pour placer dans l'espace les noeuds d'un arbre pour former un diagramme optimal [152] (voir figure 2.32). On peut distinguer deux grands groupes d'algorithmes : ceux qui mettent en valeur la nature arborescente des données (Walker) et ceux qui essaient d'utiliser au maximum l'espace disponible (placement radial, système solaire). Les approches radiales mettent moins bien en valeur l'arborescence. Une étude semble montrer que les utilisateurs préfèrent les approches radiales [66] ; elles sont cependant beaucoup moins utilisées dans les interfaces utilisateurs. Ce type d'algorithme, couplé à du *Fisheye*, a été appliqué à la visualisation de classifications médicales comme SNOMED [145] ou d'ontologie en anatomie [34].

Néanmoins, ces approches montrent vite leurs limites lorsque le nombre de noeuds devient important (de l'ordre de 100 ou 1000) : la navigation dans l'arbre devient très difficile.

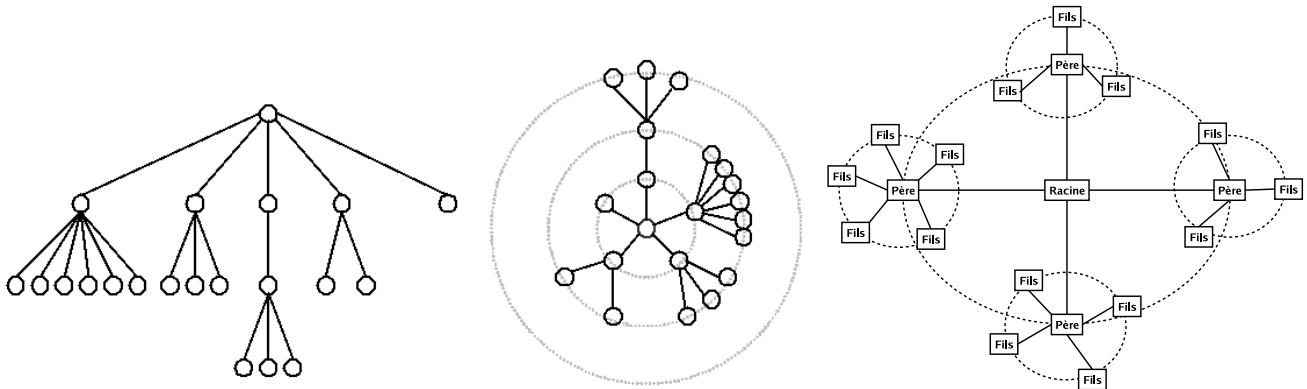


Fig. 2.32 – Exemples de diagrammes d’arbres.

À gauche : placement selon l’algorithme de Walker, au centre : placement radial (d’après [152]), à droite : placement en système solaire.

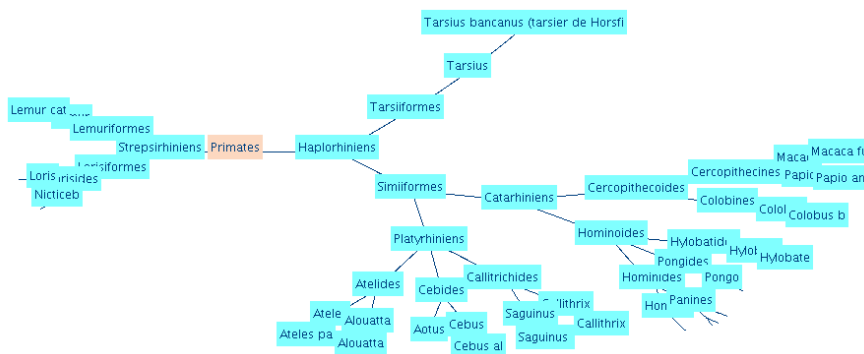


Fig. 2.33 – Exemple d’arbre hyperbolique appliqué à la classification des espèces de primates (d’après <http://www.inrp.fr/Acces/Biogeo/inxight/slsbeta/primates/primates.html>).

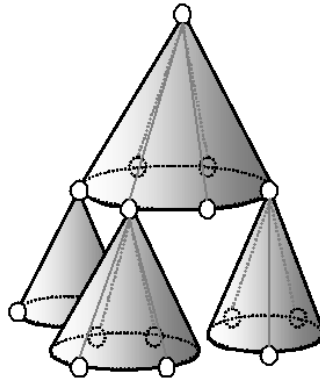


Fig. 2.34 – Exemple d’arbre conique (d’après [152]).

2.3.4.2 Arbres hyperboliques

Un arbre hyperbolique est un arbre représenté en placement radial, mais en géométrie hyperbolique. La géométrie hyperbolique provoque un *Fisheye* déformant, de sorte que les noeuds les plus éloignés du centre semblent “compressés”. L’utilisateur peut déplacer l’arbre pour positionner au centre le noeud de son choix. Cette déformation provoque un effet de *fisheye* (voir section 2.3.1) qui permet d’afficher plus d’information au niveau du noeud central, et de moins en moins lorsque l’on s’éloigne vers la périphérie.

Un inconvénient est qu’une partie non négligeable de l’espace n’est pas utilisée (dans les coins, soit $1 - \frac{\pi}{4} \simeq 21\%$ de l’espace). L’algorithme des arbres hyperboliques fait l’objet d’un brevet logiciel aux états-unis, et ne peut donc être utilisé librement.

Une démonstration d’arbre hyperbolique en Java en logiciel libre est disponible à l’adresse <http://trebolic.sourceforge.net/fr/home.htm>.

2.3.4.3 Arbres coniques (*cone tree*)

Un arbre conique est une représentation 3D d’un arbre qui permet l’affichage d’un grand nombre de noeuds (jusqu’à 1000 noeuds sur 10 niveaux de profondeur) [131]. Chaque noeud père est situé sur la pointe d’un cône, et ses fils sont placés sur le cercle qui forme la base du cône (voir figure 2.34). Les cônes sont tracés en semi-transparence, et il est possible de leur ajouter une ombre projetée pour faciliter la lecture. L’utilisateur peut faire tourner les différents cônes.

Cependant, la 3D rend les manipulations parfois délicates, et pose des problèmes d’occlusion entre les noeuds (les noeuds situés derrière sont cachés par ceux de devant). Les arbres coniques se révèlent plus efficaces pour des arbres non équilibrés, dont les branches sont suffisamment différentes.

2.3.4.4 Cartes d’arbres (*tree map*)

Les cartes d’arbres développées par Schneiderman et son équipe [77] donnent une vue d’ensemble de l’arbre. Elles utilisent la totalité de l’espace disponible; cet espace est alternativement divisé horizontalement ou verticalement, chaque case formée correspond à un noeud de premier niveau de l’arbre. Puis les cases qui contiennent des éléments fils sont elles-mêmes divisées, et ainsi de suite récursivement (voir figure 2.35). Chaque élément de l’arbre est donc représenté par une surface rectangulaire. Des variantes de l’algorithme existent, notamment pour le choix du sens de la division (horizontal ou vertical).

Les cartes d’arbres sont peu adaptées à la navigation, contrairement aux différentes représentations arborescentes vues ci-dessus [15]. Par contre elles donnent une vue d’ensemble et sont très efficaces pour le repérage de motifs récurrents dans un arbre, puisque deux branches identiques donneront deux rectangles identiques à l’écran.

Les cartes d’arbres ont principalement été appliquées à la représentation de systèmes de fichiers informatiques; d’autres applications plus récentes concernent par exemple la *Gene Ontology* [79] ou

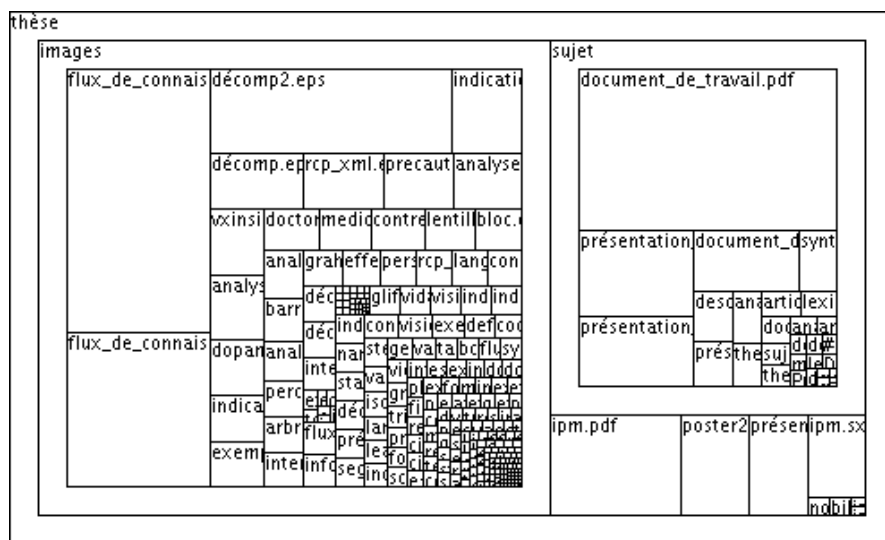


Fig. 2.35 – Exemple de carte d’arbres appliquée à la représentation d’un système de fichiers (copie d’écran de <http://treemap.sf.net>).

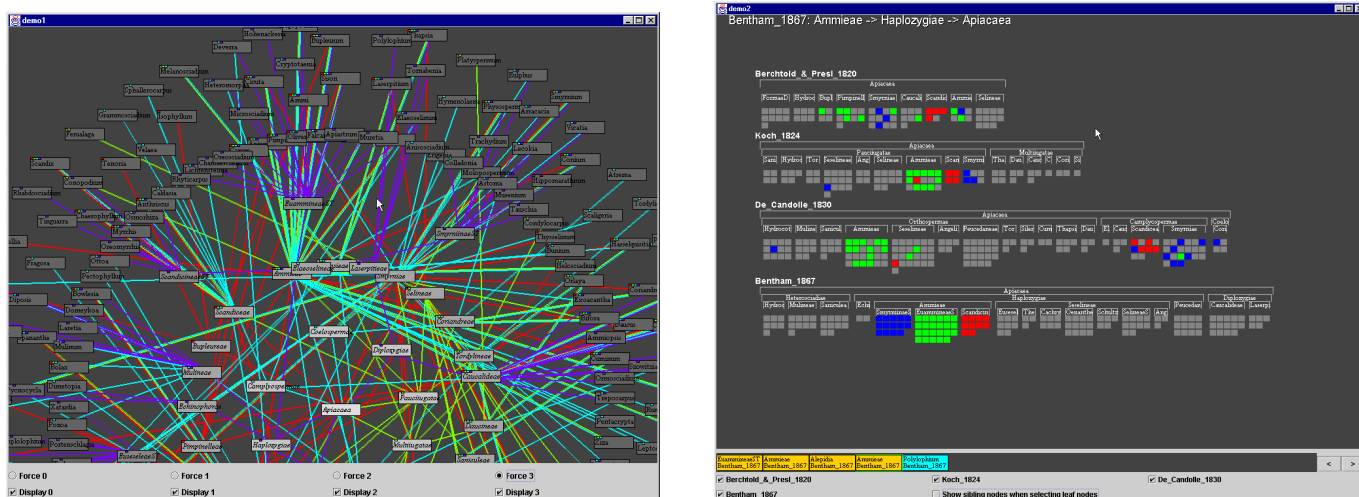


Fig. 2.36 – Représentation des arbres parallèles en taxonomie botanique par M.J. Graham [62, 61]. A gauche, représentation sous forme d’arbres superposés (une couleur par arbre) ; à droite représentation sous forme d’ensemble avec 4 classifications ; l’utilisation des couleurs permet de suivre des groupes d’espèces à travers les différentes classifications (*linking*). La seconde représentation est de loin la meilleure !

la visualisation de données médicales [26]. Voir <http://www.cs.umd.edu/hcil/treemap-history/> pour plus d’information sur les cartes d’arbres et leur évolution. Une version en 3D avec des cubes imbriqués existe aussi [71].

2.3.4.5 Arbres parallèles (*overlapping hierarchies*)

Dans sa thèse [62, 61], M.J. Graham cherche à modéliser les arbres parallèles rencontrés par exemple en taxonomie botanique : plusieurs auteurs ont défini des classifications différentes et parallèles, mais mettant en jeu les mêmes espèces et / ou groupes d’espèces. Son objectif est de représenter ces classifications parallèles pour permettre de les comparer (voir figure 2.36).

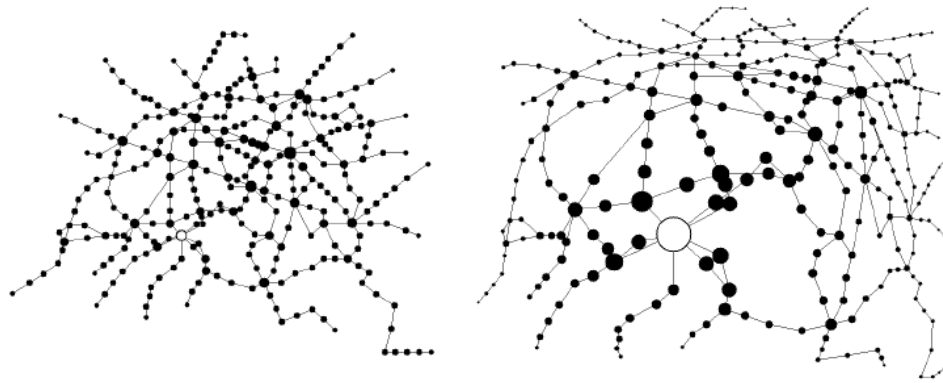


Fig. 2.37 – Exemple de *Fisheye* déformant, d'après [132].

À gauche, un réseau non déformé représentant le plan du métro parisien ; à droite, le même réseau après déformation. Le noeud central en blanc est le noeud sélectionné.

2.3.5 Méthodes de visualisation de réseaux et graphes

2.3.5.1 Diagrammes

Tout comme pour les arbres, des générateurs de diagrammes statiques peuvent être utilisés. Cependant, contrairement aux arbres, le critère d'optimisation est généralement le nombre de flèches de relation qui s'entrecroisent, ce nombre devant être le plus faible possible pour garder un diagramme lisible.

2.3.5.2 Multi-arbres

Un multi-arbre [59] est un réseau particulier où chaque noeud peut disposer de plusieurs fils et de plusieurs pères. Dans ce cas, il est possible de représenter un noeud du multi-arbre comme deux arbres distincts : un arbre des fils (*arbre des contenus*) et un arbre des pères (*arbres des contextes*). Ces deux arbres peuvent alors être représentés avec les méthodes pour les arbres vues ci-dessus.

2.3.5.3 *Fisheye* déformant

M. Sarkar et M.H. Brown ont appliqué le *Fisheye* déformant à des réseaux (figure 2.37). Cette technique peut s'ajouter aux autres techniques pour les réseaux ou les arbres.

2.3.6 Méthodes de visualisation par indices de similarité

Un indice de similarité indique la similarité entre deux objets ; plus l'indice est élevé et plus les objets sont proches. Ces méthodes s'appliquent à la visualisation d'un ensemble d'objets, lorsqu'il est possible de définir un *indice de similarité* pour chaque couple d'objets. Elles peuvent donc être appliquées aussi bien à des textes, des matrices objet-attributs, des arbres ou des réseaux.

Les méthodes à base d'indices de similarité ont surtout été appliquées à des ensembles de textes. Les indices de similarité sont obtenus après vectorisation des textes (1 mot pertinent => 1 dimension du vecteur) en calculant la distance ou l'angle entre les vecteurs de deux textes (de nombreuses variantes existent pour calculer un indice de similarité entre deux vecteurs). La vectorisation des textes conduit à des vecteurs ayant un grand nombre de dimensions (> 100) et qui ne peuvent donc pas être représentés graphiquement tel quels, d'où l'intérêt de calculer des indices de similarité, qui pourront être représentés avec les méthodes suivantes.

2.3.6.1 Terrains 3D

Il est possible de représenter sur un terrain en 3D des indices de similarité (voir figure 2.38). La hauteur du terrain dépend de la densité en objets à un endroit donné, et la distance (au sol) entre

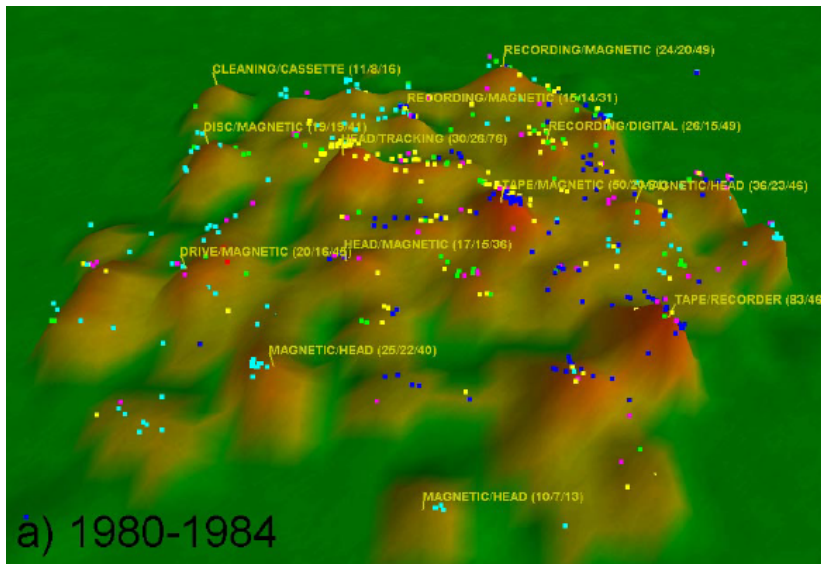


Fig. 2.38 – Terrain 3D généré par VxInsight [16] à partir d'un ensemble de textes de brevets

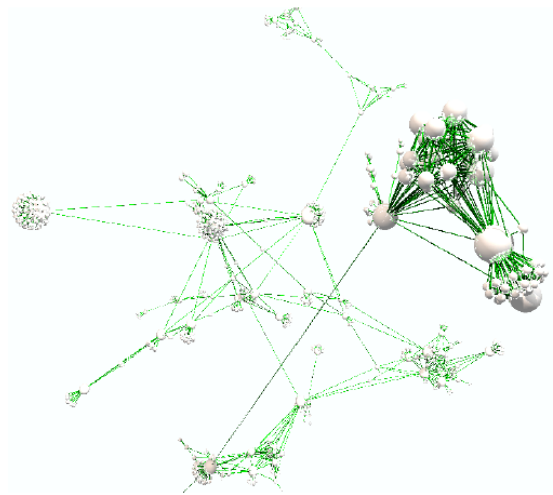
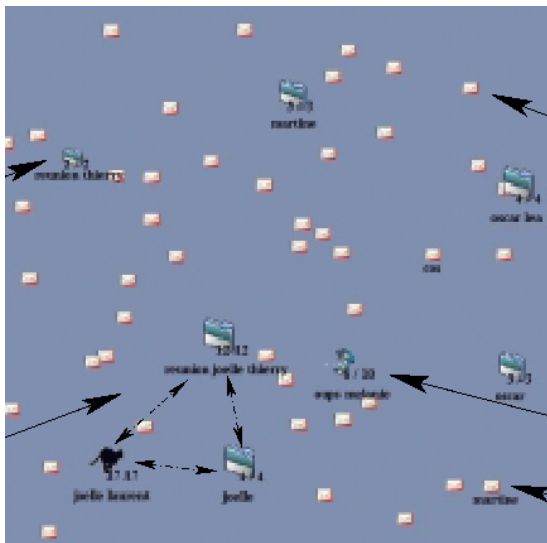


Fig. 2.39 – Exemples d'approches graphiques basées sur des forces d'attraction et de répulsion. À gauche : représentation d'un ensemble de messages e-mails en 2D par LEA [128] ; à droite : représentation d'un réseau d'hyperliens en 3D par Narcissus [70].

deux objets représente leur similitude (les objets similaires sont proches les uns des autres et forment donc des pics). Les deux dimensions horizontales (X, Z) n'ont pas de signification exacte. Ce mode de représentation fait appel à la loi de proximité de la Gestalt (voir section 2.1.2).

Les objets sont d'abord placés de manière aléatoire, et des forces d'attraction / répulsion sont définies entre chaque couple d'objets selon les règles suivantes : (a) deux objets similaires (indice élevé) s'attirent, (b) deux objets très proches se repoussent (cette seconde règle évite que tous les objets ne s'agglomèrent entre eux). La position finale des objets est obtenue après stabilisation du système de forces, généralement après un grand nombre d'itérations ou après des recuit-simulés (augmentation puis diminution de l'énergie totale du système, afin de permettre le franchissement des minimums locaux).

2.3.6.2 Système multi-agents

Des systèmes multi-agents ont été proposés pour représenter des indices de similarité, chaque élément étant considéré comme un agent et interagissant avec les autres éléments en fonction de la similarité entre les deux éléments et de la distance physique qui les sépare.

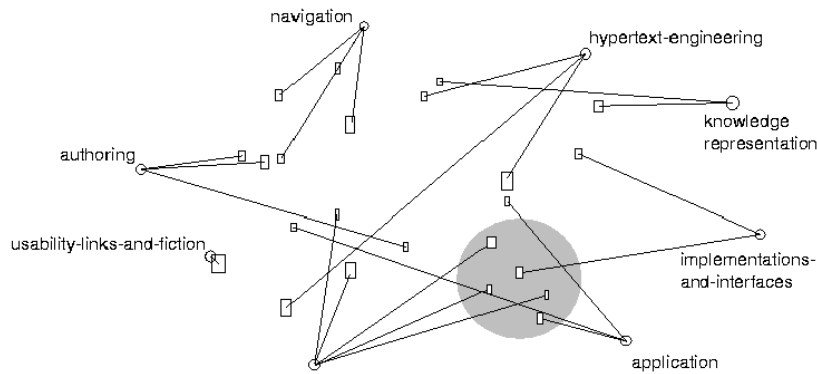


Fig. 2.40 – VIBE [104] (d'après [68]).

Les POI sont représentés par des ronds et un texte, les objets (ici des documents) par des rectangles de taille proportionnelle à leur intérêt.

Dans sa thèse, V. Renault [128] propose une approche multi-agents inspirée de l'étude des sociétés d'insectes (éthologie). A chaque objet correspond un agent mobile soumis à des forces d'attraction / répulsion dépendant en particulier de la similarité entre les agents. A la différence d'un terrain 3D, l'utilisateur voit le déplacement des objets et peut influencer sur ce déplacement, par exemple en déplaçant manuellement un objet.

Cette méthode a été appliquée à la visualisation de boîtes aux lettres électroniques (figure 2.39 à gauche), chaque objet étant un message e-mail. Les objets très proches se regroupent pour former des répertoires. L'utilisateur définit des mots clefs qui sont utilisés pour le calcul des indices de similarité. Cependant la mobilité des agents peut poser problème et rendre difficile la recherche d'un objet précis.

Les forces d'attraction-répulsion peuvent aussi être définies dans un espace en 3D, contrairement au cas précédent qui se situe dans un plan, ou au terrain 3D où la 3^{ème} dimension a une signification différente. C'est ce que propose Narcissus [70] pour cartographier le Web (figure 2.39 à droite).

2.3.6.3 Points d'intérêt

Une autre méthode est celle des points d'intérêt (*Point Of Interest*, POI) présentée par A. Olsen *et al.* [104]. L'utilisateur définit un certain nombre de *points d'intérêt* (par exemple des mots clefs) et les positionne sur un plan. Les indices de similarité ne sont pas calculés pour chaque couple d'objets mais pour chaque couple (POI, objet) ; l'absence d'interaction objet-objet permet d'éviter les systèmes de forces d'attraction / répulsion souvent difficiles et coûteux à résoudre.

Chaque POI correspond à une dimension différente ; cette méthode permet donc de projeter un espace à n dimensions sur un plan ou un espace. La projection peut donc masquer certaines différences entre les objets, et deux objets différents peuvent se retrouver au même endroit. Dans certains cas, il est possible de les séparer en déplaçant les POI.

La méthode des POI a été utilisée en 2D (VIBE [104], figure 2.40) et en 3D (Lyberworld).

Chacune des visualisations présentées ci-dessus peut être plus ou moins abstraite : par exemple les terrains 3D peuvent avoir l'apparence de surfaces 3D ou bien de véritables terrains (couleur verte / marron, présence de niveau d'eau). Nous allons maintenant étudier de plus près l'utilisation de ces métaphores.

2.3.7 Métaphores

Certaines techniques ci-dessus utilisent déjà la 3^{ème} dimension de manière métaphorique (comme les terrains 3D section 2.3.6.1 ou les arbres coniques section 2.3.4.3). C.R. Dos Santos [40] a proposé

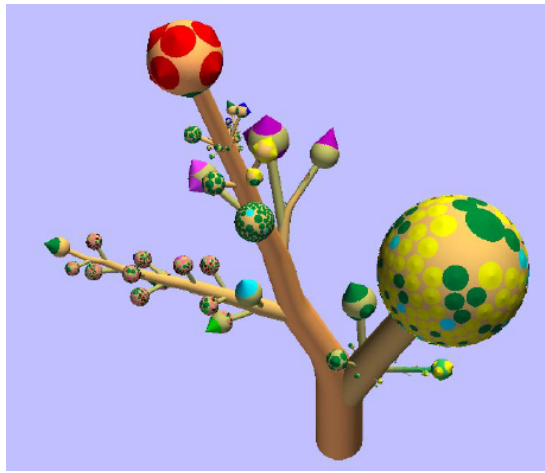


Fig. 2.41 – Exemple d'arbre botanique (d'après [80]).

une revue de l'utilisation des métaphores dans les visualisation 3D, et G. Sherson [138] a étudié l'utilisation des métaphores sur les sites Internet. Nous allons présenter ici quelques exemples.

2.3.7.1 Arbre botanique

Kleiberg et son équipe [80] proposent de représenter des arbres sous forme d'arbres (voir figure 2.41), d'une manière proche des arbres coniques. Chaque branche de l'arbre peut donner naissance a plusieurs autres branches, et les éléments de la branche qui n'ont pas d'éléments fils (parfois appelés feuilles) sont représentés sur un fruit au bout de la branche ; le fruit porte un cône par élément. Par ailleurs, cette forme de représentation peut donner lieu à des images très esthétiques.

2.3.7.2 Villes

T. Panas *et al.* [108] proposent de représenter sous forme de villes des programmes informatiques, dans le but d'aider au développement et à l'ingénierie logicielle par une représentation globale d'un logiciel en cours de réalisation. Chaque paquetage (ensemble de classes) du code source est une ville en 3 dimensions. Les différentes villes sont reliées entre elles par des voies d'eau (liaison unidirectionnelle) ou des routes (liaison bidirectionnelle) ; des bateaux ou des voitures indiquent les échanges et les appels qui ont lieu entre les paquetages.

A l'intérieur d'une ville, chaque classe du paquetage est représentée par un immeuble, dont la taille est proportionnelle à celle de la classe. Des informations supplémentaires figurent sur les immeubles : les immeubles en feu sont les classes qui ont demandées le plus de travail, les immeubles en ruine celles qui doivent être revues, en jaune les classes sur lesquelles les développeurs travaillent,...

2.3.8 Conclusion sur la visualisation d'information

Le tableau 2.2 récapitule les différentes techniques de visualisation d'information que nous avons vues et leurs points forts. Les classifications habituelles reposent sur la structure des informations visualisées. Cependant, en ce qui concerne les connaissances sur le médicament, ces classifications sont peu utiles. En effet, selon la manière dont on les considère, les connaissances sur le médicament peuvent avoir différentes structures :

Texte les Résumés des Caractéristiques Produits et les Guide de Bonnes Pratiques se présentent sous forme de texte libre (Vidal électronique).

Indices de similarité il est possible d'extraire des indices de similarité à partir des textes des RCP, ou à partir d'un modèle de connaissance structuré.

Graphique 2D, 3D le médicament est un objet à n dimensions, quantitatives (prix, date de mise sur le marché,...) ou qualitatives (peut-il être prescrit chez l'enfant,...).

type	technique	<i>Fisheye</i>	cas propices
texte	repliement	F	texte hiérarchisé
texte	lentille de document	D	grande quantité de texte, texte en pages
texte	mur fuyant	D	grande quantité de texte
texte	<i>greeking</i>	D	très grande quantité de texte, mais illisible
nD	graphiques		<7 dimensions dont 2 quantitatives au moins (X, Y)
nD	coordonnées parallèles		valeurs quantitatives
nD	glyphe		recherche d'objets similaires
nD	lentille de tableau	F	matrice objet-attribut, <1000 lignes à l'écran
nD	réduction réciproque de liste	F	valeurs qualitatives
nD	requête dynamique	F	données rapidement requêttables (<100 ms)
(tous)	lentille magique	F/D	requête restreinte à une région de l'espace
arbre	diagramme		<100 noeuds
arbre	système solaire		<3-4 niveaux
arbre	arbre hyperbolique	D	<1000 noeuds (dont 50 étiquetés)
arbre	arbre conique	D	arbre non balancé, <1000 noeuds
arbre	carte d'arbre		recherche de motif, <1000 noeuds
arbre	arbre parallèle		plusieurs arbres avec les mêmes objets, <1000 noeuds
réseau	diagramme		<100 noeuds
réseau	multi-arbre		réseau particulier, < 100 noeuds
réseau	fisheye déformant	D	< 200-500 noeuds?
(tous)	terrain 3D		recherche d'objets similaires
(tous)	points d'intérêt		

Tab. 2.2 – Caractéristiques des différentes techniques de visualisation d'information.

Pour chaque technique est indiqué le type de données, l'utilisation éventuelle du *Fisheye* (F : filtrant, D : déformant) et les cas propices.

Matrice objet-attribut médicament-attribut (Dopamine)

Arbres par exemple classement des médicaments par classes thérapeutiques, ou classification chimique (Vidal électronique).

Réseau par exemple un réseau reliant les médicaments et les pathologies entre eux, par des liens d'indication, contre-indication, interaction médicamenteuse, effet indésirable,...

En effet, lors d'une utilisation clinique, **la structure des informations médicales n'intéresse que peu le médecin**. Le choix de la méthode de visualisation devra donc plutôt être fait en fonction de la tâche que le médecin souhaite réaliser. Pour cela, je propose **une classification des méthodes de visualisation d'information s'appuyant sur l'usage auquel l'outil de visualisation est destiné** (cette classification s'inspire de celle que j'avais proposée lors d'une conférence [86]). Le tableau 2.3 présente les six usages que nous avons distingués et indique quelles sont les techniques de visualisation les plus adaptées. Remarquons que cette classification diffère de la classification par structure habituelle : par exemple, les différentes techniques de visualisation d'arbres ont été dispersées dans les usages "vue d'ensemble", "recherche par similarité" et "navigation". Le tableau 2.4 liste des exemples d'application dans le domaine médical. Les applications sont classées selon le public à qui elles s'adressent (patient ou professionnel) et selon l'usage concerné. Pour le patient, seul les usages "vue d'ensemble" et "navigation" semblent réellement intéressants.

Cette classification distingue les approches destinées au patient et celles destinées aux médecins, puis parmi celles-ci les approches permettant la navigation au sein d'un document ou d'une liste, la comparaison entre plusieurs éléments similaires, la représentation de vues d'ensemble et la recherche de similarité.

usage	exemple	techniques de visualisation
vue d'ensemble : donner une vue d'ensemble sur un sujet	Quels sont les principales propriétés du Stilnox [®] ?	<i>Fisheye</i> , carte d'arbre, <i>greekling</i> , lentille de document, mur fuyant
recherche : rechercher un élément dont seuls certains aspects sont connus	Quelle classe thérapeutique d'anti-hypertenseur prescrire pour mon patient qui est âgé et diabétique ?	lentille de tableau, requête dynamique, réduction réciproque de liste, point d'intérêt
recherche par similarité : rechercher des éléments proches d'un élément donné	Quel médicament a des propriétés similaires au Stilnox [®] ?	glyphes, indices de similarité, terrains 3D, cartes d'arbre
navigation : retrouver un élément précis, déjà connu	Je recherche la monographie du Stilnox [®] .	arbre hyperbolique, arbre conique, diagramme, <i>Fisheye</i> , repliement
comparaison : comparer un petit nombre d'éléments entre eux	Parmi la classe des diurétiques de l'anse, lequel est le plus adapté à mon patient ?	lentilles de tableau, coordonnées parallèles, graphiques, point d'intérêt
vérification : vérifier la présence ou l'absence d'un élément	Ce médicament est-il contre-indiqué en cas d'asthme ?	<i>Fisheye</i> , repliement

Tab. 2.3 – Les différents usages des approches graphiques.

public	usage	exemples d'application
patient	<u>vue d'ensemble</u>	document pédagogique expliquant les causes et les effets d'une pathologie, et le mode d'action des traitements associés. résumé des principales propriétés du médicament, simplifié à destination du patient. vue d'ensemble du plan de prise.
	recherche	(application peu envisageable car le patient ne connaît pas les propriétés des pathologies ou des médicaments)
	recherche de similarité	-
	<u>navigation</u>	naviguer dans une classification graphique simplifiée pour permettre au patient d'entrer ses pathologies et symptômes sans utiliser le langage médical.
	comparaison	-
	vérification	vérifier si le traitement a bien été pris.
professionnel de santé	<u>vue d'ensemble</u>	vue d'ensemble des patients d'un hôpital. résumer le dossier d'un patient. résumer les principales propriétés d'un médicament.
	<u>recherche</u>	rechercher quelles pathologies peuvent causer tel symptôme. rechercher quelle classe thérapeutique prescrire pour un patient donné. rechercher les patients ayant telle pathologie. rechercher des documents ou des articles traitant de tel sujet.
	recherche de similarité	rechercher des profils de patients similaires dans un hôpital. lorsque le médicament que le médecin voulait prescrire est contre-indiqué, rechercher un médicament similaire.
	<u>navigation</u>	naviguer au sein d'une classification des médicaments (ATC par exemple) pour chercher une classe thérapeutique donnée.
	<u>comparaison</u>	comparer différentes pathologies ayant des tableaux cliniques proches (aide au diagnostic différentiel). comparer les différentes classes thérapeutiques traitant une même pathologie. comparer des profils de patients. comparer des résultats d'examen. comparer plusieurs plans de prises possibles.
	<u>vérification</u>	vérifier pour un médicament donné une contre-indication précise, correspondant par exemple à une pathologie du patient. vérifier si tel médicament peut causer tel effet indésirable (par exemple dans un centre de pharmaco-vigilance)

Tab. 2.4 – Exemples d'applications des différents usages, pour le patient et pour les professionnels de santé. Les usages les plus intéressants sont soulignés.

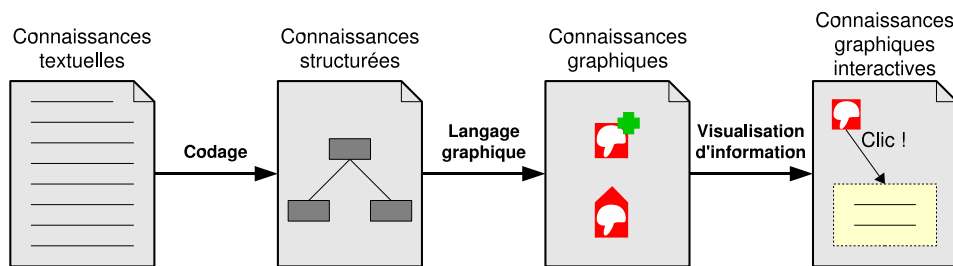


Fig. 2.42 – Les 4 formes de la connaissance : textuelle, structurée, graphique et graphique interactive.

Les métaphores peuvent rendre une interface plus intuitive et plus attrayante. Cependant elles peuvent aussi ajouter un bruit visuel inutile. L'aspect attractif de ce genre de visualisation est intéressant pour attirer des utilisateurs et notamment des patients, cependant elle n'est pas forcément adaptée aux professionnels de santé qui risquent d'être noyés dans des détails non pertinents et de ne pas avoir des informations assez précises.

2.4 Conclusion : quelle approche graphique pour les connaissances sur le médicament ?

Dans ce chapitre, nous avons d'abord présenté la sémiotique et les différents modèles de la vision humaine. Ensuite, nous avons étudié deux approches de visualisation. Les langages graphiques représentent des informations ou des connaissances à l'aide de combinaison de pictogrammes, de couleurs,... plutôt qu'avec des lettres ou des mots. La visualisation d'information permet de représenter, souvent de manière interactive, la structure des informations.

La visualisation d'information permet de représenter des concepts généraux liés à la structure des informations, par exemple l'arbre d'une classification des pathologies ou les chiffres issus d'une série d'analyses biologiques. Par contre, elle ne parvient pas à représenter des concepts qualitatifs spécifiques à ces informations, comme la nature des pathologies ou les différentes analyses biologiques. Elle peut représenter les résultats d'un test biologique chez 100 patients, mais ne peut pas représenter les résultats de 100 tests différents chez un même patient. En conséquence, la visualisation d'information n'est pas adaptée pour représenter des connaissances (voir section 3.1 pour la différence entre information et connaissance, et 8.4.4 à propos de la visualisation de connaissance). C'est pourquoi nous concentrerons l'essentiel de nos efforts sur les langages graphiques.

En revanche, une fois qu'un langage graphique a été utilisé pour traduire graphiquement les connaissances sur le médicament, il est possible d'appliquer des techniques de visualisation d'information sur les icônes obtenues. Les deux approches présentées, langage graphique et visualisation d'information, sont donc complémentaires (figure 2.42). Les trois questions auxquelles nous nous intéressons (voir section 1.2) font principalement appel aux usages de vue d'ensemble et de vérification, et dans une moindre mesure de navigation (au sein d'un texte). Le tableau 2.3 suggère pour ces trois usages l'utilisation du *Fisheye*.

La démarche de conception que nous suivrons correspond à celle présentée dans la figure 2.42 : dans un premier temps nous étudierons les connaissances sur le médicament afin de modéliser ce que nous cherchons à représenter. Ensuite, nous construirons un langage graphique pour représenter ces connaissances et nous lui appliquerons des techniques de visualisation d'information. Enfin, le langage graphique et les outils de visualisation mis au point feront l'objet d'une évaluation dans des conditions contrôlées.

Chapitre 3

État de l’art des connaissances sur le médicament

Les connaissances sur le médicament sont de quatre types : pharmaceutiques, cliniques, pharmacologiques et administratives ; dans le cadre de notre travail, ce sont principalement les connaissances cliniques qui nous intéressent. Les sources de connaissances sur le médicament se divisent en deux branches : les sources orientées médicaments (principalement les monographies ou RCP), et les sources orientées patients (les guides de bonnes pratiques).

Les sources originelles et officielles sont en texte libre, et sont donc difficilement exploitables informatiquement. C’est pourquoi des modèles ont été mis au point pour structurer les connaissances sur le médicament et concevoir des bases de connaissance.

Dans ce chapitre, nous définirons tout d’abord le mot “connaissance”, puis nous verrons les différentes sources de connaissances sur le médicament, ainsi que les modèles développés pour ces sources et les classifications médicales mises à profit pour cela. Enfin, nous nous intéresserons à la littérature sur les besoins en connaissances sur le médicament exprimés par les différents acteurs.

3.1 Définitions de la connaissance

Les notions de données, d’information et de connaissance sont sujettes à de nombreux débats ; les définitions suivantes sont données uniquement pour clarifier la suite de l’exposé et n’ont pas de vocation “universelle”. F. Nake [102] propose une définition de ces 3 termes à partir de la sémiotique et de la triade peircienne : R est la Représentation, O l’Objet représenté et I l’Interprétation (voir section 2.1).

Les **données** correspondent à la dimension *syntactique*. Une donnée est un *signe*, par ex : “la glycémie du patient A est de 1,5g/l”. La représentation (R) est interprétée comme étant elle même (dans l’exemple, $R = O = I =$ glycémie de 1,5g/l).

L’**information** correspond à la dimension *sémantique*. Une information est une donnée à laquelle on associe un sens, par ex. : “la glycémie du patient A est de 1,5g/l donc le patient A est diabétique”¹. Elle est composée d’une représentation et d’un objet (R - O) ; la représentation représente l’objet qui est interprété comme étant lui-même (dans l’exemple, $R =$ glycémie de 1,5g/l, $I = O =$ le patient A est diabétique).

La **connaissance** correspond à la dimension *pragmatique*. Une connaissance est une information réutilisable dans un autre contexte ; cette notion est subjective et dépend de l’utilisateur. Par ex. : “si la glycémie est de 1,5g/l, le patient est diabétique” ; cette règle est générale et réutilisable sur n’importe quel patient, un exemple d’interprétation subjective est “donc le patient B hyperglycémique qui est dans mon service est aussi diabétique”. La connaissance est une triade complète ((R - O) - I) : R représente O qui est interprétée pour former une interprétation I, différente de O car subjective (dans l’exemple, $R =$ glycémie de 1,5g/l, $O = A$ est diabétique, $I = B$ est diabétique).

¹En pratique médicale, plusieurs mesures de la glycémie sont nécessaires pour réaliser le diagnostic du diabète.

3.2 Les monographies des médicaments

3.2.1 Les monographies sous forme textuelle

Les monographies des médicaments reposent sur les RCP (Résumé des Caractéristiques Produits). Un RCP regroupe l'ensemble des connaissances cliniques, pharmaceutiques, pharmacologiques et administratives sur un médicament donné, et sert de document de référence. Les RCP sont écrits par les industries pharmaceutiques, puis validés par l'AFSSAPS (Agence Française de Sécurité Sanitaire des Produits de Santé), qui les met ensuite à disposition des éditeurs de bases de connaissances médicamenteuses : Vidal, Resip (banque Claude Bernard), GieSIPS (Thériaque). Certains éditeurs ont tendance à interpréter les RCP pour essayer de corriger les erreurs alors que d'autres s'en tiennent à des RCP bruts. Les informations suivantes sont contenues dans les RCP :

- nom de marque,
- composition qualitative et quantitative,
- forme pharmaceutique,
- aspects cliniques :
 - indications : dans quelles situations prescrire le médicament,
 - posologie et méthode d'administration : à quelle posologie prescrire le médicament, et comment l'administrer,
 - contre-indications : dans quelles situations le médicament ne doit pas être prescrit (cette section peut reprendre des éléments d'autres sections, comme "mise en garde", "interactions médicamenteuses" ou "grossesse et allaitement", et l'on distingue les contre-indications relative des contre-indications absolues),
 - précautions d'emploi et mises en garde : les mises en garde indiquent un risque iatrogénique² possible, ou bien une action thérapeutique à effectuer (ex. : diminuer la posologie, effectuer une surveillance,...) ou au contraire à ne pas faire (ex. : ne pas arrêter le traitement brusquement) ; les précaution d'emploi sont des mises en garde qui ne s'appliquent que chez certains patients (ex. : chez les personnes âgées),
 - interactions médicamenteuses : avec quels autres médicaments le médicament interagit,
 - grossesse et allaitement : si le médicament peut être prescrit en cas de grossesse, ou en cas d'allaitement (cette section contient aussi des informations sur le passage du médicament dans le lait maternel, et le résultat d'éventuelles expérimentations animales),
 - effets indésirables : les effets défavorables et de nature fortuite survenant au cours de la prise normale du médicament à dose habituelle,
 - surdosage : les signes de surdosage, et ce qu'il convient de faire en cas de surdosage,
- propriétés pharmacologiques : pharmacodynamie, pharmacocinétique, données de sécurité pré-clinique,
- aspects pharmaceutiques : liste des excipients, incompatibilités, durée de stabilité, conditions de stockage, nature des conditionnements, instructions pour la manipulation,
- identité du propriétaire de l'AMM (Autorisation de Mise sur le Marché),
- numéro d'AMM,
- date de la première autorisation ou de renouvellement d'AMM,
- date de révision du texte.

Les monographies sont la stricte reproduction des RCP auxquels sont ajoutés le prix, le coût du traitement journalier, et l'inscription sur les listes des substances vénéneuses. Dans le cadre de notre travail, **ce sont principalement les aspects cliniques qui nous intéressent.**

3.2.2 Modèles structurés des monographies

De nombreux modèles ont été proposés pour structurer les connaissances des monographies. Les différentes bases de connaissances (Vidal, Banque Claude Bernard, Thériaque,...) possèdent chacune

²iatrogénique = causé par le traitement médical

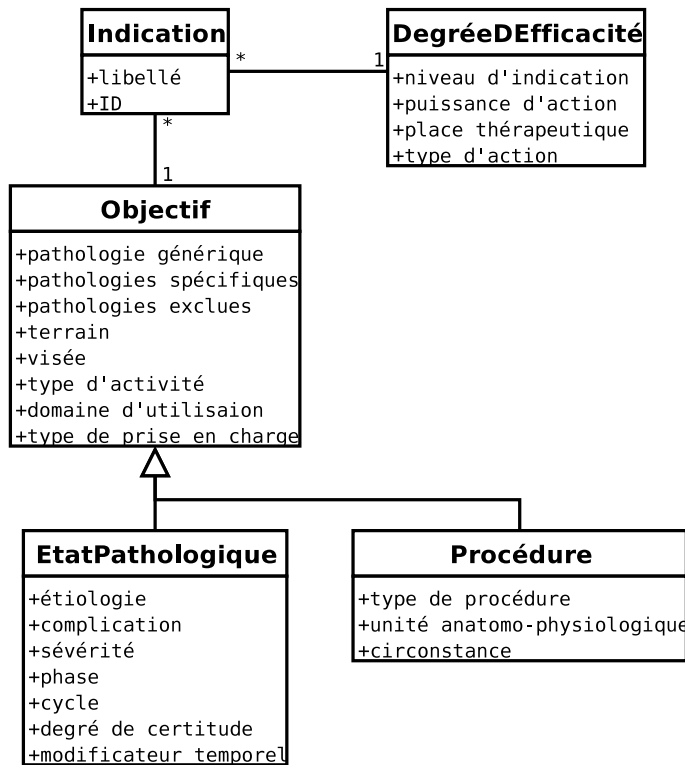


Fig. 3.1 – Modèle UML des indications des RCP, d’après [43, 45, 154]

leurs propres modèles, plus ou moins détaillés. Nous mentionneront tout d’abord des modèles XML à but administratif, puis nous verrons quelques modèles issus de laboratoires de recherche portant sur des sections cliniques du RCP. La thèse de C. Duclos propose un état de l’art complet de ces modèles [43].

3.2.2.1 Modèles XML

Ces modèles s’appuient sur XML ou SGML [72]. Ces initiatives s’intéressent à l’aspect administratif et à l’identification du médicament, plutôt qu’à ses propriétés cliniques. Elles sont donc d’un intérêt limité dans notre contexte.

eCTD (*electronic common technical document*, <http://esubmission.eudra.org/ectd.html>) est un standard d’échange électronique du dossier de demande d’AMM, commun à l’Europe, au Japon et aux USA. eCTD repose sur XML, cependant celui-ci n’est utilisé que comme “enveloppe”; les connaissances cliniques du dossier ne sont pas structurées.

L’EMA (*European Agency for the Evaluation of Medicinal Product*) étudie PIM (*Product Information Management*), un système pour permettre la soumission en ligne des informations concernant les médicaments par les industriels. PIM repose sur un fichier unique en XML, qui regroupe les informations du RCP, de la notice patient et de l’emballage, dans les différentes langues européennes.

SPC-DTD [158] est une DTD destinée à l’écriture de RCP (SPC en anglais). Cependant cette DTD ne structure que les différentes sections du RCP, et ne descend pas plus bas (seules des balises de mise en forme (emphase,...) sont utilisées à l’intérieur des sections).

3.2.2.2 Indications

C. Duclos [43, 45, 154] propose un modèle objet pour représenter les indications des RCP (figure 3.1). Ce modèle a été construit après une analyse lexico-sémantique des sections “indications” des RCP, et a été évalué en terme de variabilité et de couverture. 3 niveaux d’indication ont été définis, par ordre décroissant : “indiqué pour...” (efficacité cliniquement démontrée), “proposé pour...” (activité pharmacologiquement démontrée, mais pas cliniquement) et “utilisé pour...” (activité non démontrée).

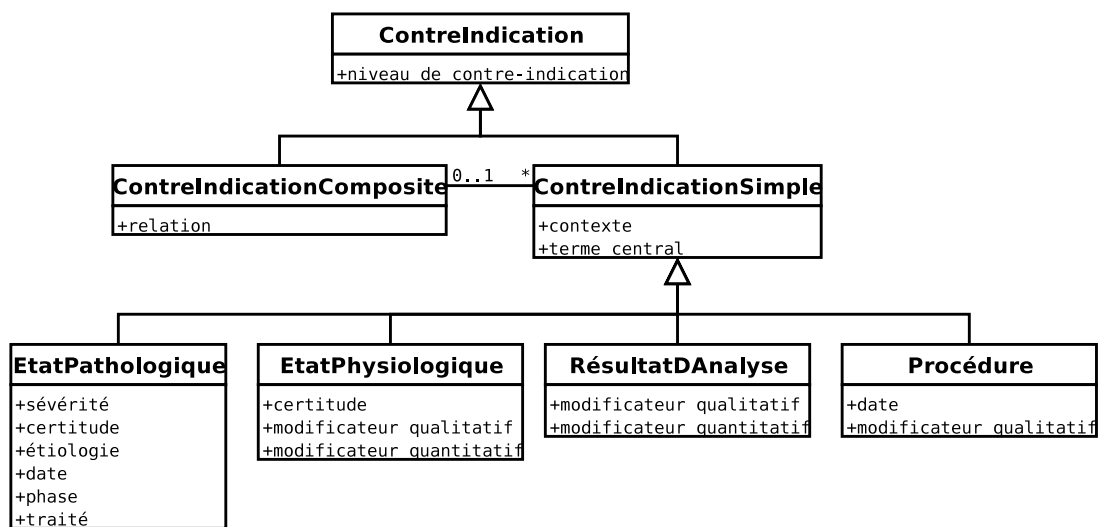


Fig. 3.2 – Modèle UML des contre-indications des RCP, d’après [93]

3.2.2.3 Contre-indications

J.H. Liu *et al.* ont proposé un modèle objet pour représenter les contre-indications [93] (figure 3.2). La notion de contexte (attribut de la classe *ContreIndicationSimple*) permet de représenter la relation entre le patient et la personne concernée par l’état pathologique / physiologique / ... Il existe 5 contextes :

- l’état actuel du patient,
- l’historique du patient (par exemple antécédents thrombo-emboliques),
- l’historique maternel (par exemple mère VIH+),
- l’historique familial (par exemple antécédents familiaux de glaucome),
- l’état actuel du partenaire (par exemple partenaire susceptible de tomber enceinte).

3.2.2.4 Précautions d’emploi et mises en garde

Le groupe de travail PE-MEG (A. Venot, C. Duclos, 2003, rapport interne au laboratoire) a défini les concepts que l’on retrouve dans les précautions d’emploi (PE) et les mises en garde (MG), suite à l’analyse de 10000 précautions d’emploi-mises en garde différentes. Les conditions qui définissent les patients concernés peuvent porter sur un état pathologique, un état physiologique, un mode de vie (tabac, alcool, sport, voyage à l’étranger,...), un résultat de laboratoire ou l’historique thérapeutique des prescriptions médicamenteuses (traitement par le médicament porteur de la précaution d’emploi-mises en garde, ou autre traitement), procédures ou règles hygiéno-diététiques.

3.2.2.5 Interactions médicamenteuses

Les interactions médicamenteuses interviennent entre un médicament et un ou plusieurs autres éléments parmi d’autres médicaments, des résultats de laboratoire, des aliments (figure 3.3). En général, les interactions médicamenteuses mettant en jeu plus de deux médicaments sont systématiquement négligées, à cause du trop petit nombre de sujets concernés (qui ne permet pas d’obtenir une preuve statistique). Pour chaque interaction, il est possible de préciser le niveau de gravité, le mécanisme d’action, l’effet, le délai entre les prises et les conduites à tenir en cas d’interaction.

3.2.2.6 Effets indésirables

Un effet indésirable est un “effet défavorable et de nature fortuite lorsqu’il survient au cours de la prise normale d’un médicament aux doses habituelles” (définition de l’OMS). Il peut être décrit

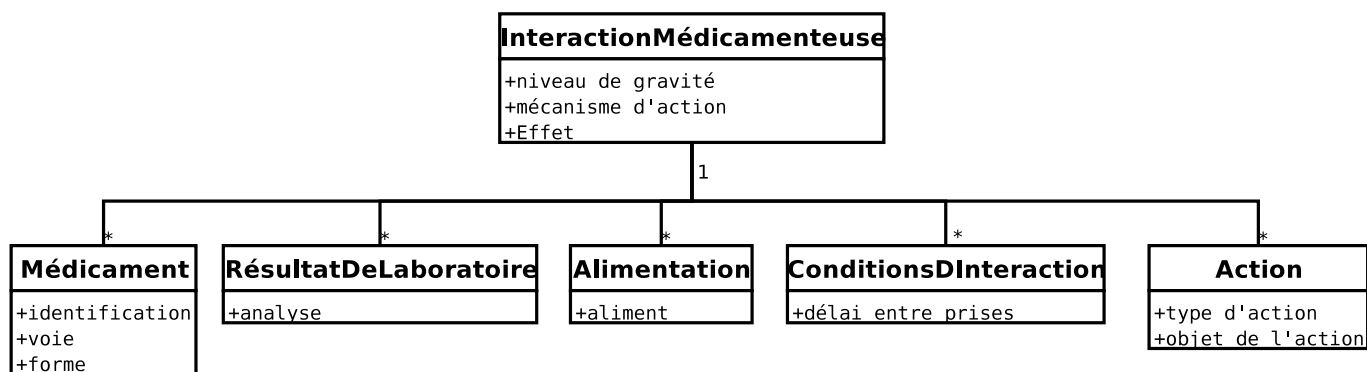


Fig. 3.3 – Modèle UML des interactions médicamenteuses des RCP, d’après [43]

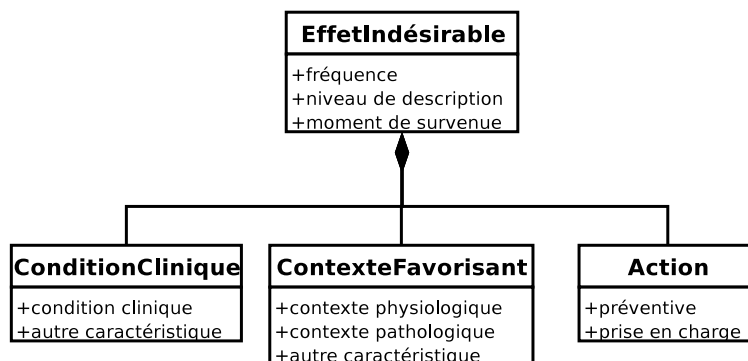


Fig. 3.4 – Modèle UML des effets indésirables des RCP, d’après [43]

(figure 3.4) par une situation clinique se produisant à une certaine fréquence et qui est favorisée par certains contextes physiologiques ou pathologiques. Des actions peuvent être entreprises pour éviter ou corriger l’effet indésirable.

3.2.2.7 Pharmacocinétique

Un modèle très complet de la pharmacocinétique des RCP a été réalisé par C. Duclos [46, 43]. Ce modèle découpe la pharmacocinétique en 4 parties (voir version simplifiée en figure 3.5) :

Processus pharmacocinétique les *données explicatives* décrivent l’évolution du médicament dans l’organisation au cours des 4 processus : absorption, distribution, métabolisme, élimination. Chaque processus comprend une ou plusieurs réactions, qui comprend elle même un ou plusieurs mécanismes. Une réaction transforme des substances initiales en substances finales (transformation chimique ou déplacement d’un compartiment à un autre). Un ensemble de mesures temporelles qualitatives ou quantitatives peut aussi être donné ; ces mesures peuvent évoluer sur plusieurs phases.

Protocole expérimental il décrit les conditions d’obtention des données pharmacocinétiques : protocole d’administration (administration, posologie), protocole de mesure et population concernée (espèce, états physiopathologiques).

Modèle mathématique il décrit le modèle cinétique compartimental des processus pharmacocinétiques et ses différents paramètres (demi-vie,...).

Facteurs de variation ils décrivent les modifications apportées par des changements de protocole expérimental aux mesures, aux paramètres ou aux processus.

3.3 Bases de connaissances médicamenteuses

Les connaissances contenues dans les monographies sont ensuite informatisées et structurées en France dans différentes bases de connaissances, comme Vidal, la Banque Claude Bernard ou Thé-

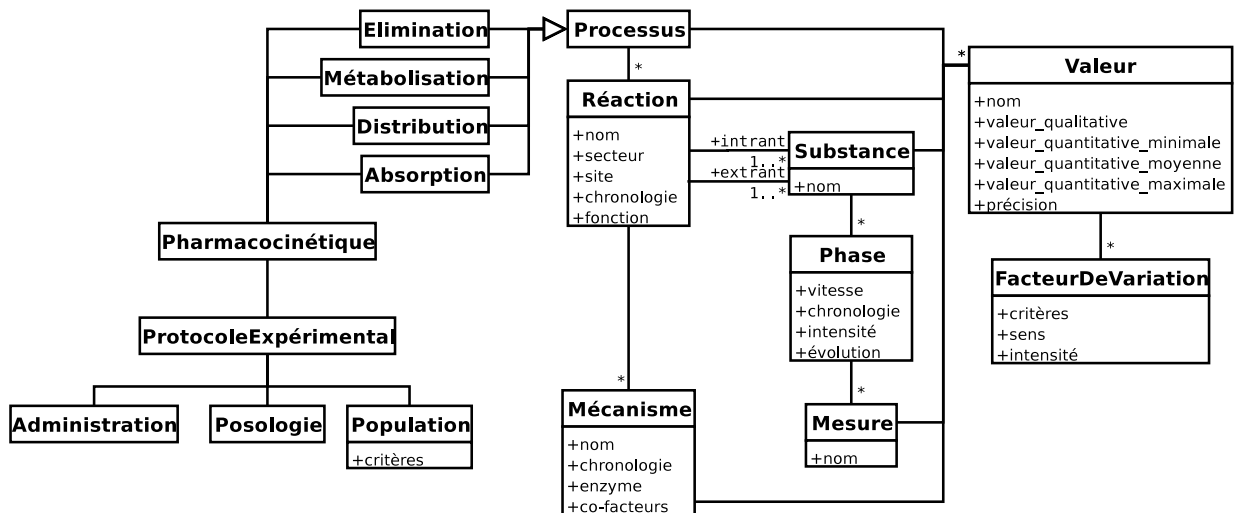


Fig. 3.5 – Modèle UML de la pharmacocinétique, version simplifiée d'après [43].

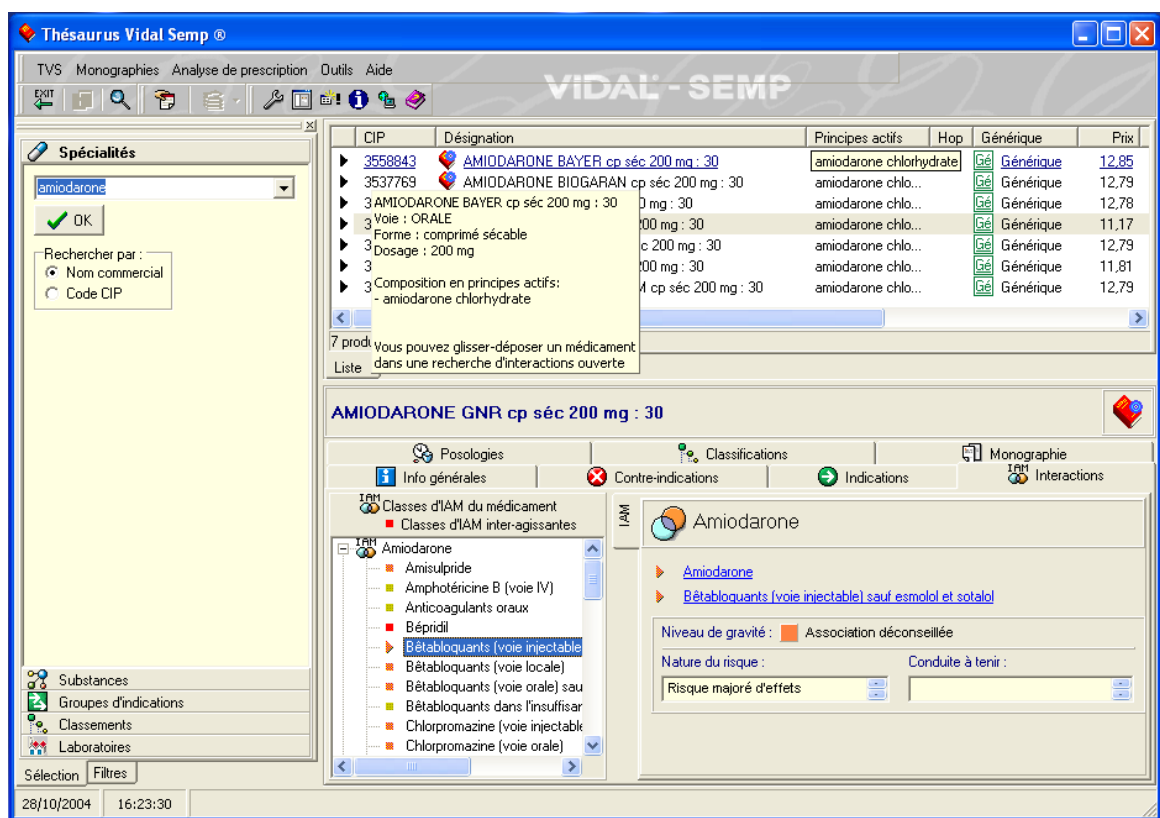


Fig. 3.6 – Copie d'écran de la monographie de l'amiodarone dans le Vidal électronique.

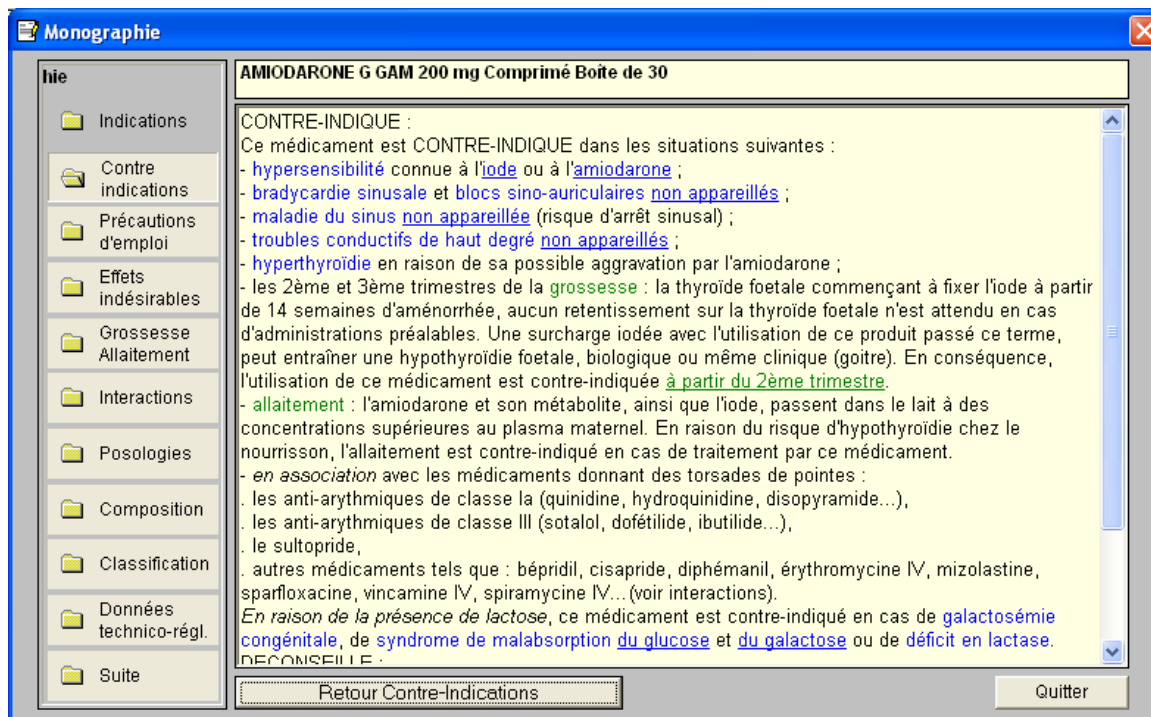


Fig. 3.7 – Copie d’écran de la monographie de l’amiodarone dans la Banque Claude Bernard.

riatique. Galen Drug Ontology [162, 82] est un exemple de base de connaissance fortement structurée, basée sur les logiques de descriptions.

En pratique libérale, deux outils sont utilisés : le **Vidal électronique** (environ deux tiers du marché, figure 3.6) et la **Banque Claude Bernard** (BCB, environ un tiers du marché, figure 3.7). Dans ces deux bases de connaissances médicamenteuses, l’aspect graphique se limite à des icônes pour les différents menus et des couleurs pour les différents niveaux de contre-indications dans le Vidal électronique, et à des mises en forme du texte (souligné ou en couleur) dans la Banque Claude Bernard.

Le mode de consultation des bases de connaissances sur le médicament a peu évolué malgré l’informatisation : un dictionnaire électronique reste avant tout un dictionnaire ! D’après une enquête sur les bases de données médicamenteuses électroniques réalisée par FULMEDICO (Fédération des Utilisateurs de Logiciel Médicaux et Communicants) [58], “Actuellement, on a [...] l’impression d’être en présence d’une informatisation des bases de données médicamenteuses ‘papier’ ”. Les utilisateurs évoquent notamment la quantité excessive des connaissances et le temps nécessaire pour trouver les connaissances pertinentes.

3.4 Guide de bonnes pratiques

3.4.1 Les guides de bonnes pratiques sous forme textuelle

Les Guides de Bonnes Pratiques (GBP) contiennent des recommandations à suivre pour la prise en charge d’une catégorie de patients (par exemple atteints d’une pathologie donnée) ; contrairement aux RCP, les GBP sont “orientés patients”. Les GBP sont rédigés par des institutions (AFSSAPS, HAS en France) ou des sociétés savantes (par exemple l’Association Médicale Canadienne a réalisé un GBP sur l’hypertension [2]), et mis à disposition des professionnels de santé et des éditeurs de logiciel.

Un des problèmes rencontrés avec les GBP est la présence de “trous” dans ces guides : certains patients n’entrent dans aucune catégorie, et d’autres entrent dans plusieurs catégories. Pour certains guides, ces cas sont très fréquents et l’utilisation du GBP par le médecin est alors très frustrante : de nombreux médecins se plaignent sur les forums ou au cours d’enquêtes [58] que les applications

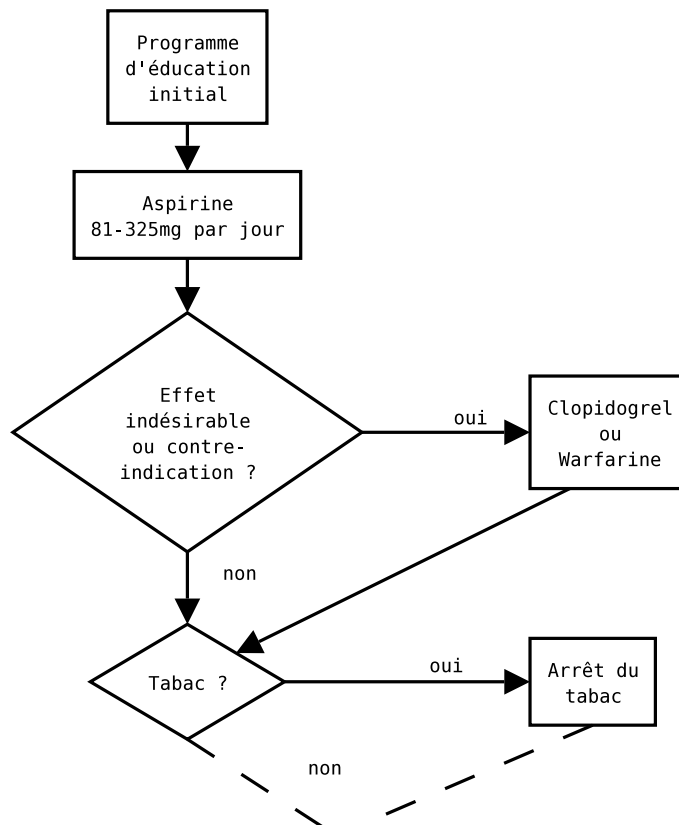


Fig. 3.8 – Exemple d’arbre de décision extrait d’un guide de bonnes pratiques sur l’angine de poitrine (d’après <http://www.openclinical.org/>).

informatiques à base de GBP sont pour eux une perte de temps.

3.4.2 Modèles structurés des GBP

Un grand nombre de modèles ont été développés spécifiquement pour les GBP ; ces modèles présentent une grande variété d’approches. Des arbres de décision (voir figure 3.8) sont utilisés par GLIF3 (*GuideLine Interchange Format*) [111, 110], PROforma [146] et GLARE [64] pour représenter les GBP. Un formalisme à base de règles “si... alors...” est utilisé par le langage XML GEM [139]. Le *workflow* (ou *careflow*) a été utilisé dans GUIDE [35] pour gérer des processus parallèles ou séquentiels, des événements (*trigger*), des contraintes ou des limites de temps. Les GBP peuvent aussi être représentés sous forme de scénarios, correspondant à des “patients théoriques” (par exemple un patient hypertendu traité par une monothérapie), et d’actions, qui permettent de passer d’un scénario à un autre (par exemple prescrire un second médicament à ce patient fera passer dans le scénario bithérapie) ; c’est le cas dans PRODIGY [119, 78] et EON [148]. Enfin, ASBRU [137] se base essentiellement sur la description de plan temporel en utilisant un langage spécifique. Voir <http://www.openclinical.org/gmmintro.html> pour plus d’informations et de références sur l’informatisation des GBP et les différents langages de représentation des GBP.

Cependant la plupart des modèles réalisés pour les GBP ne peuvent pas être appliqués aux RCP. C’est notamment le cas de tous les modèles impliquant un processus de raisonnement (arbres de décision, règles “si... alors...”, réseaux de Pétri).

3.5 Classifications médicales

Les différents modèles de structuration des connaissances sur le médicament utilisent tous des classifications médicales, des terminologies ou des vocabulaires contrôlés. Ceux-ci permettent de standardiser les noms ou identifiants des pathologies, des médicaments,... et évitent les problèmes de

synonymie, d'ambiguïté ou de fautes d'orthographe que l'on rencontre avec le texte libre. Cependant, ces classifications n'ont pas toujours une précision suffisante pour représenter la totalité des informations présentes dans les textes de référence (RCP ou GBP). Dans cette section, nous allons présenter plusieurs classifications couramment appliquées aux connaissances sur le médicament ; la thèse de C. Duclos effectue une présentation plus complète [43].

3.5.1 CIM10

La CIM10 (Classification Internationale des Maladies version 10, <http://www.icd10.ch>, maintenue par l'OMS) a été conçue à l'origine pour coder les causes de décès en épidémiologie. Elle est utilisée aujourd'hui largement en dehors de ce cadre originel, et est disponible en plusieurs langues (dont le français). La classification se présente sous la forme d'un arbre, cependant elle contient aussi des relations inter-branches, reliant par exemple une complication clinique et sa cause. Elle contient environ 20000 termes classés en 21 chapitres.

Les maladies sont réparties en catégories selon une étiologie, un système ou un processus morbide commun. Outre les maladies, la CIM10 contient aussi les symptômes, les modes de vie, les résultats d'analyses anormaux, les blessures et empoisonnements, les causes externes de morbidité et de mortalité, les facteurs influençant l'état de santé, ainsi que certaines procédures ou consultations médicales.

3.5.2 SNOMED RT

La SNOMED (*Systematized Nomenclature of Medicine*, <http://snomed.org>) est une terminologie médicale multi-axiale ; elle propose des codes aussi bien pour les pathologies que les médicaments, les procédures médicales,... SNOMED RT (*Reference Terminology*) ajoute des relations sémantiques entre les termes, par l'utilisation de logique de description.

3.5.3 MedDRA

MedDRA (*Medical Dictionary for Drug Regulatory Affairs*) est une terminologie pour décrire les aspects réglementaires du développement du médicament (dossiers d'essais cliniques, d'enregistrement d'AMM et de pharmacovigilance,...). Elle propose entre autres des termes pour décrire les symptômes, les maladies, les diagnostics, les procédures, l'historique du patient,... Cependant seuls les termes pertinents pour les aspects réglementaires sont présents. La classification est multi-axiale. La capacité de MedDRA à représenter les concepts des RCP a été évaluée à environ 60% pour les indications, les formes et l'administration du médicament, 75% pour les contre-indications, les précautions d'emploi et les surdosages, et 90% pour les effets indésirables [17].

3.5.4 ATC

La classification ATC (*Anatomical Therapeutic Chemical classification of drugs*) est une classification des principes actifs des médicaments développée par le *Nordic Council of Medicine*, puis reprise par l'OMS. Elle est traduite en plusieurs langues, dont le français. La classification est arborescente, et les différents niveaux représentent successivement les systèmes anatomiques, les classes thérapeutiques et les classes chimiques (chacun s'étalant éventuellement sur plusieurs sous-niveaux).

Un même principe actif peut se voir attribuer plusieurs codes ATC, s'il appartient à plusieurs classes, par exemple parce qu'il possède plusieurs indications ou modes d'administration.

3.5.5 CCAM

La CCAM (Classification Commune des Actes Médicaux, <http://www.le-pmsi.org/ccamv0b/>) est une nomenclature des actes médicaux à vocation exhaustive et destinée au codage et à la tarification des actes. Cette classification est réalisée par l'Assurance Maladie et le PERNNS (Pôle

d'Expertise et de Référence National des Nomenclatures de Santé), et validée par l'HAS. La CCAM est destinée à remplacer la NGAP (Nomenclature Général des Actes Professionnels) et le CdAM (Catalogue des Actes Médicaux), incomplets et contradictoires.

La CCAM regroupe environ 7100 actes différents, classés de manière hiérarchique selon 4 axes : action, topographie, voie et technique utilisée, par exemple : biopsie / du rein / par voie transcutanée / sans guidage. Les deux derniers axes sont optionnels. Pour les termes anatomiques, la nomenclature anatomique internationale francisée a été reprise. Un outil de recherche en logiciel libre, Népal (<http://medecinelibre.nuxeo.org/nepal/>), permet de parcourir la CCAM.

3.5.6 UMLS

UMLS (*Unified Medical Language System*) est un projet de la NLM (*National Library of Medicine*) qui cherche à établir des liens entre différentes sources de connaissance utilisant des codages différents. UMLS comprend un métathésaurus et un réseau sémantique. Le métathésaurus regroupe tous les termes issus d'une centaine de classifications comme la CIM10, et établit des relations d'équivalence, d'héritage, d'appartenance,... entre eux. Chaque terme est associé à un type sémantique, et le réseau sémantique relie ces types sémantiques par des relations sémantiques.

3.6 Besoins en connaissance sur le médicament

Nous avons vu que les connaissances médicamenteuses regroupent plusieurs catégories de connaissance (voir section 3.2). Ces connaissances intéressent de nombreuses personnes, des professionnels de santé aux patients. Les besoins exprimés par ces utilisateurs doivent être pris en compte lors de la réalisation de notre approche graphique, notamment afin de déterminer quelles connaissances seront visualisées.

3.6.1 Besoins des professionnels de santé

Les connaissances médicamenteuses sont utiles à la fois au médecin, au pharmacien et à l'infirmière. Elles peuvent être utilisées dans trois situations différentes :

En face du patient durant la consultation (médecin), la délivrance du médicament (pharmacien) ou l'administration du médicament (infirmière).

Lorsque le patient n'est pas présent, pour approfondir un point spécifique.

Lors de la formation initiale ou continue.

La première situation, en face du patient, est celle où le manque de temps est le plus important à cause de la présence du patient. C'est donc la situation idéale pour une approche graphique, afin de réduire le volume des connaissances et d'accélérer leur accès.

Lors de la prescription, le médecin procède en plusieurs étapes : choix du traitement, vérification des contre-indications, établissement de la posologie et du plan de prise, conseils au patient et suivi (si nécessaire). Le pharmacien suit des étapes similaires : choix du médicament dans le cas de génériques, vérification des contre-indications, établissement du plan de prise (si le médecin ne l'a pas fait), conseils au patient et suivi (si nécessaire).

L'étude Formmel 2001 [41] a permis d'évaluer la satisfaction et les attentes des médecins vis-à-vis des logiciels et sites Internet médicaux. Les médecins utilisent pour la plupart les modules de détection des contre-indications et des interactions médicamenteuses et d'alarme des logiciels d'aide à la prescription. Les bases de connaissances médicamenteuses (type "Vidal électronique") sont utilisées principalement pour les contre-indications, les effets secondaires, les indications et la posologie (par ordre décroissant de fréquence d'utilisation). Elles sont peu utilisées pour rechercher quel médicament prescrire ; et un certain nombre de médecins considèrent que les modules de ce type (guides de bonnes pratiques, fiches de transparence...) sont une perte de temps.

Le pharmacien quand à lui est un “fournisseur de médicaments” mais aussi un “fournisseur de connaissances” sur les médicaments [3]. Les connaissances qui intéressent plus particulièrement le pharmacien sont les interactions médicamenteuses, les contre-indications et les informations pour la réalisation d’un plan de prise (pharmaco-cinétique, chrono-pharmacologie,...), ainsi que toutes les informations demandées par le patient (prise du médicament, effet indésirable, conseil au patient,..., voir ci-dessous). Les préparations magistrales sont de plus en plus rares ; les besoins d’information associés concernent surtout la stabilité des préparations (notamment pour les suspensions orales) [33].

3.6.2 Besoins des patients

Le patient n’est pas l’utilisateur direct de nos outils de visualisation, cependant les besoins exprimés par le patient doivent être pris en compte, afin que le médecin ou le pharmacien aient accès aux connaissances nécessaires à l’éducation du patient. Évaluer les besoins du patient est une tâche difficile ; les études se font avec des questionnaires de sortie, des enquêtes périodiques massives, des interviews plus approfondies, ou par le biais d’associations de patients [105]. Le patient est très demandeur d’information sur son traitement et il s’intéresse principalement aux informations suivantes [83, 38, 39, 11] (par ordre décroissant d’intérêt) :

- les effets indésirables,
- les actions du médicament et la raison pour laquelle il doit le prendre (pathologie traitée, action préventive, curative ou symptomatique),
- les modifications à apporter aux habitudes de vie du patient,
- la manière de prendre le médicament,
- la composition et la nature du médicament,
- les interactions avec les autres médicaments pris au long cours,
- la conduite à tenir si les symptômes persistent ou disparaissent,
- l’efficacité du traitement.

Cependant, **les médecins n’accordent pas la même importance aux différentes informations à donner aux patients** [11]. En particulier, les effets indésirables sont souvent négligés, certains médecins craignant (souvent à tort) qu’informer le patient sur les effets indésirables ne diminue l’observance, ou n’augmente la fréquence de ces effets (“effet nocebo”). En première position, les médecins placent les interactions avec les médicaments en vente libre.

L’HAS a publié un ensemble de recommandations sur l’information au patient [147].

3.7 Conclusion

Nous avons défini la connaissance, puis nous avons vu les deux types de connaissances sur le médicament : connaissances orientées médicament (RCP) et orientées patient (GBP). De nombreux modèles ont été proposés pour ces connaissances, cependant il n’existe pas de modèle suffisamment général pour représenter à la fois les deux types de connaissances sur le médicament. De plus, ces modèles sont très précis, trop pour servir de base à la conception d’un langage graphique, forcément moins précis que les connaissances textuelles. Nous avons aussi étudié les besoins en connaissance sur le médicament, et mis en évidence la différence entre les besoins des professionnels et ceux des patients.

Nous pouvons construire **une classification des outils de visualisation des connaissances médicales** destinés aux professionnels de santé, en nous appuyant sur les étapes de la prescription ou de la délivrance du médicament. Chacune de ces étapes (choix du traitement, vérification des contre-indications, établissement de la posologie et du plan de prise, conseil au patient et suivi) utilise des connaissances médicamenteuses différentes. Nous proposons de classer les outils de visualisation en fonction de l’étape à laquelle ils se destinent :

Outil de *navigation* permettant au médecin de choisir un médicament ou une classe thérapeutique

adapté à un patient donné parmi l'ensemble des médicaments ou des classes ; dans ce cas on s'intéressera principalement aux indications et aux contre-indications.

Outil de *comparaison* permettant au médecin ou au pharmacien (dans le cas d'une substitution par un générique) de choisir un médicament parmi plusieurs médicaments proches (une classe thérapeutique par exemple) ; dans ce cas on s'intéressera surtout aux différences entre les médicaments, en terme d'optimisation thérapeutique (minimiser les effets indésirables, améliorer le plan de prise,...) et de coût de traitement.

Outil de *vérification* montrant sous forme de résumé graphique les propriétés essentielles d'un médicament sur un seul écran. Cet outil permet au médecin ou au pharmacien de vérifier que la prescription ou la délivrance du médicament est possible. Il fait intervenir essentiellement les contre-indications, les interactions médicamenteuses et les précautions d'emploi-mises en garde. Deux utilisations sont possibles : soit l'utilisateur regarde l'ensemble des contre-indications, des interactions médicamenteuses et des précautions d'emploi-mises en garde par "acquis de conscience", soit il recherche une contre-indication précise correspondant à une pathologie ou un traitement du patient (asthme par exemple).

La pharmacocinétique peut intervenir à ce niveau (par exemple l'association de deux traitements fortement liés aux protéines plasmatiques risque de poser problème) ; elle est pourtant très peu utilisée lors de la pratique médicale courante. Il serait cependant intéressant de résumer par des icônes les principales caractéristiques pharmaco-cinétiques d'une spécialité : inducteur / inhibiteur enzymatique, élimination rénal / hépatique, métabolisation, fixation aux protéines plasmatiques, demi-vie importante.

Outil de *posologie et plan de prise* sous forme de résumé graphique, pour le médecin et le pharmacien.

Outil d'*aide à l'éducation du patient* montrant sous forme de résumé graphique les conseils à donner au patient ; ce type d'outil doit prendre en compte les besoins exprimés par les patients, et en particulier les effets indésirables. Il est difficile de "faire le tri" entre les très nombreux effets indésirables indiqués dans le RCP. Cet outil doit présenter une vue d'ensemble des effets indésirables en mettant en évidence les plus importants, c'est à dire les plus fréquents, les plus graves, et ceux qui donnent lieu à des conseils à donner au patient (par exemple éviter le soleil).

Outil de *suivi* montrant sous forme de résumé graphique les propriétés d'un médicament utiles lors du suivi du traitement : surveillance, problème d'observance, que faire si le traitement est inefficace, effets indésirables, surdosage. Lors du suivi, si un patient se plaint d'un symptôme donné (par exemple de toux), cet outil doit aider à déterminer s'il s'agit d'un effet indésirable, d'un signe de surdosage ou d'un symptôme sans rapport avec le médicament.

Outil de *lecture* semi-graphique (mélange texte + image) facilitant la lecture du texte par l'utilisation de couleurs, de mises en forme, d'icônes, et de techniques de visualisation. Cet outil est utilisé en complément des précédents, lorsque les autres outils n'ont pas apporté suffisamment d'information.

L'analyse des besoins a montré un manque d'intérêt de la part des médecins pour les outils de type navigation ou comparaison, c'est-à-dire permettant de choisir quel médicament prescrire. Les outils de posologies et de plans de prise correspondent à un besoin et se prêtent bien à des représentations graphiques, cependant il existe déjà des initiatives dans ce domaine [149]. Nous avons donc choisi de nous intéresser aux outils de vérification, d'éducation du patient et de lecture. Ces outils font appel aux aspects cliniques suivants du RCP : contre-indications, interactions médicamenteuses, effets indésirables et précautions d'emploi / mises en garde.

Nous allons maintenant nous intéresser à la conception d'un modèle des connaissances sur le médicament adapté à un langage graphique.

Chapitre 4

Modélisation des connaissances sur le médicament

4.1 Introduction et objectifs

La construction d'un langage graphique doit s'appuyer sur une analyse approfondie des connaissances à représenter. Les modèles des connaissances sur le médicament existants (voir chapitre 3) ont été développés dans le but de concevoir des bases de connaissance sur le médicament, et ils ont un niveau de précision important. Or un langage graphique comme VCM ne cherche pas à représenter la totalité des informations disponibles sur un concept donné, mais un minimum d'information permettant d'identifier ce concept. Par exemple il n'est pas nécessaire de préciser que le diabète de type 2 est une maladie chronique pour identifier cette pathologie.

L'objectif de ce chapitre est donc d'étudier les connaissances sur le médicament afin de savoir précisément ce que nous voulons représenter, et de concevoir un modèle des connaissances sur le médicament suffisamment simple pour servir de base pour des langages graphiques comme VCM.

4.2 Matériels et méthodes

Notre modèle doit pouvoir représenter les principaux *concepts* et les *phrases* construites à l'aide de ces concepts. Un *concept* est une "construction mentale" issue de l'effort abstraction de l'esprit, qui essaie de représenter une unité de sens [127]. En médecine, les concepts correspondent généralement à la vision des experts du domaine, par exemple des pathologies (ex. : le concept d'hypertension) ou des médicaments (ex. : le concept d'antidiabétique),... Pour concevoir notre modèle, nous devons donc déterminer :

les types de concepts médicaux (pathologies, médicaments,...) mis en jeu par les connaissances sur le médicament,

les attributs utilisés pour identifier ces concepts (localisation anatomique ou fonctionnelle d'une pathologie, activité pharmacologique d'un médicament,...) ainsi que leurs **valeurs** possibles (cardiaque, pulmonaire,... pour la localisation par exemple),

les relations entre ces concepts ; nous nous sommes principalement intéressé aux relations d'héritage *est-un*, car nous verrons que ce sont les seules qui peuvent être représentées graphiquement de manière immédiate (voir section 5.4.1),

la structure des phrases médicales mettant en jeu ces concepts (par exemple, le concept de pathologie est utilisé dans les phrases exprimant des contre-indications).

Les langages graphiques étant peu précis (voir section 1.2), des choix doivent être fait pour garder un modèle simple. Lors de ces choix, en particulier pour déterminer les attributs que nous avons retenus, trois critères ont été pris en compte : l'importance médicale, la fréquence d'apparition dans les connaissances sur le médicament, et la facilité de représentation graphique. La méthodologie

de conception s'est appuyée sur trois points : les connaissances d'experts, les modèles existants et l'analyse des connaissances contenues dans un corpus de RCP par des méthodes de Traitement Automatique du Langage naturel (TAL).

4.2.1 Connaissances d'expert du domaine

Cette première étape a été réalisée à partir de mes connaissances d'expert (en tant que pharmacien) et de celles d'autres membres du laboratoire, médecins ou pharmaciens. Leur prise en compte est nécessaire, car **pour que le langage VCM soit facile à apprendre pour les professionnels de santé, il doit faire appel aux concepts et attributs qu'ils ont l'habitude de manipuler dans leur pratique clinique**, et qu'ils ont rencontrés lors de leur formation. Par exemple notons l'importance associée à l'axe anatomico-fonctionnel, qui sert en particulier à définir les spécialités médicales (cardiologie, pneumologie,...), aux différentes étiologies infectieuses (bactérienne, virale, mycosique et parasitaire) ou aux classes d'âge.

Les connaissances d'expert ont aussi été utilisées pour déterminer **la facilité de représentation graphique** et **l'importance** des différents attributs envisagés. Certains attributs ne peuvent pas être représentés de manière graphique, et n'ont donc pas pu être retenus. C'est en particulier le cas des dénominations : Dénomination Commune Internationale (DCI), nom de spécialité, nom ou famille chimique pour un médicament, nom de la personne qui a étudié une maladie (ex. : maladie de Parkinson), nom de substance (ex. : dosage de la calcémie). Pour ces derniers, il existe des symboles (Na pour le sodium, Ca pour le calcium,...), cependant ceux-ci sont de nature textuelle et nous avons préféré, au moins dans un premier temps, nous en tenir à des représentations strictement non textuelles.

L'importance des attributs a été évaluée à partir de trois critères : l'importance médicale intrinsèque, la pertinence pour répondre à l'une des trois questions : (a) quelles sont les principales propriétés du médicament X ? (b) le médicament X est-il contre-indiqué chez mon patient Y ? (c) que dois-je dire sur le médicament X à mon patient Y ? (voir section 1.2), et la redondance. En effet, certains attributs sont redondants avec d'autres, par exemple il n'y a que peu de cibles pharmacologiques pour lesquelles il existe à la fois des médicaments activateurs et inhibiteurs, donc l'attribut "activité sur la cible pharmacologique" est redondant par rapport à l'attribut "cible pharmacologique" : si un médicament agit sur l'enzyme de conversion, il n'est pas nécessaire de préciser que le médicament est un inhibiteur. Dans ce cas, nous considérerons l'attribut "activité sur la cible pharmacologique" comme peu important.

4.2.2 Modèles existants

Une bibliographie des modèles existants dans la littérature a été donnée dans les sections 3.2 et 3.4.2. Nous avons vu en introduction de ce chapitre que même si ces modèles n'étaient pas adaptés aux langages graphiques, ils peuvent cependant nous aider. **Les concepts et attributs rencontrés dans les différents modèles sont similaires** et se recoupent ; par exemple comparer la représentation des états pathologiques dans les modèles d'indication et de contre-indication des sections 3.2.2.2 et 3.2.2.3. Certains attributs, comme l'étiologie ou la sévérité, semblent faire consensus. Les modèles proposés pour les différentes sections du RCP (indications, contre-indications, interactions médicamenteuses, précautions d'emploi-mises en garde, effets indésirables) donnent aussi une idée de la structure des phrases correspondantes et des concepts qu'elles contiennent.

Les modèles proposés pour les GBP ont exploré une grande variété d'approches. Parmi celles-ci, **les approches à base de règles "si... alors..." ou de scénarios sont les plus adaptées à un langage graphique pour représenter les connaissances sur le médicament**, afin d'exprimer des conseils ou des interdictions s'appliquant à certaines catégories de patients. Les arbres de décisions se prêtent bien à des représentations graphiques, mais les RCP n'ont pas cette structure.

Les classifications médicales présentées à la section 3.5 définissent toutes des relations *est-un*, entre les concepts d'un même type : les pathologies pour la CIM10, les médicaments pour l'ATC,... Ces

relations pourront être reprises dans notre modèle. De plus, elles pourront aussi être utilisées pour remplacer par des concepts plus généraux les concepts trop précis que VCM ne peut représenter : par exemple, le concept “myocardite aiguë” sera représenté comme “pathologie cardiaque”.

4.2.3 Analyse d’un corpus de RCP

L’analyse d’un corpus de RCP avait pour objectif de fournir des éléments chiffrés pour la conception du modèle, tels que les fréquences d’apparition des différents attributs et de leurs valeurs. Cette analyse expérimentale a été réalisée en utilisant des outils de TAL. Nous verrons d’abord les textes et outils utilisés, puis la méthode appliquée.

4.2.3.1 Matériels

Le choix du corpus s’est porté sur les RCP, car ceux-ci sont les documents de référence sur le médicament. Nous n’avons pas travaillé sur la totalité des RCP (6000 documents environ) pour des raisons pratiques. Notre analyse a donc porté sur les 278 RCP des spécialités disponibles en France et présents sur la 13^{ème} liste des médicaments essentiels de l’OMS (http://www.gujhealth.gov.in/Medi_servi/WHO%20Formulary/mf_misc/medicines_alphabetical_order.asp.htm). Ce groupe de RCP a été décomposé en 4 corpus, comprenant respectivement les sections contre-indications, effets indésirables, interactions médicamenteuses et précautions d’emploi-mises en garde de chaque RCP.

Les trois outils de TAL suivant ont été utilisés :

Cordial (voir <http://www.synapse-fr.com/>) est un logiciel de correction d’orthographe que nous avons utilisé comme lemmatiseur (il détermine la forme canonique des mots : le masculin singulier pour les noms et adjectifs, l’infinitif pour les verbes) et étiqueteur grammatical (il détermine la catégorie grammaticale des différents mots d’une phrase : nom, verbe, adjectif,...).

Lexter (voir http://www.atala.org/article.php3?id_article=140) est un analyseur syntaxique réalisé par Didier Bourigault à EDF ; il permet d’extraire à partir d’un corpus l’ensemble des syntagmes nominaux.

TreeTagger ([134], voir <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html>) est lui aussi un lemmatiseur et un étiqueteur grammatical multilingue (dont français).

Nous avons utilisé des lemmatiseurs différents (Cordial et TreeTagger) dans nos deux analyses. Le choix du lemmatiseur a été réalisé uniquement pour des raisons matérielles (machine Windows pour Cordial, puis machine Linux / Unix pour TreeTagger).

4.2.3.2 Analyse des attributs, leurs fréquences et de leurs valeurs

Afin de déterminer les attributs utilisés pour identifier les principaux concepts dans les RCP, ainsi que leurs fréquences d’utilisation et leurs différentes valeurs, nous avons mis au point une méthode d’analyse lexico-sémantique. En utilisant Lexter, nous avons obtenu la liste des candidats-termes, c’est-à-dire des noms (ex : hypertension) ou des syntagmes nominaux (ex : hypertension artérielle sévère), dans leur forme canonique (masculin singulier). Ensuite, une analyse sémantique a été réalisée de manière semi-automatique sur ces candidats-termes, afin de déterminer quels sont ceux qui appartiennent à un type de concept médical donné (par exemple les pathologies, ou les médicaments), puis les attributs qui le définissent. Voici l’algorithme utilisé (figure 4.1) :

1. Soit CT1 la liste des candidats-termes générés par Lexter. À partir de CT1, extraire la liste de tous les noms et adjectifs, sans doublon, que nous appellerons W1.
2. Marquer manuellement dans la liste W1 quels sont les mots marquant le type de concept qui nous intéresse (par exemple “trouble”, “maladie”, “cardiopathie”, “hypertension”,... si l’on s’intéresse aux pathologies).

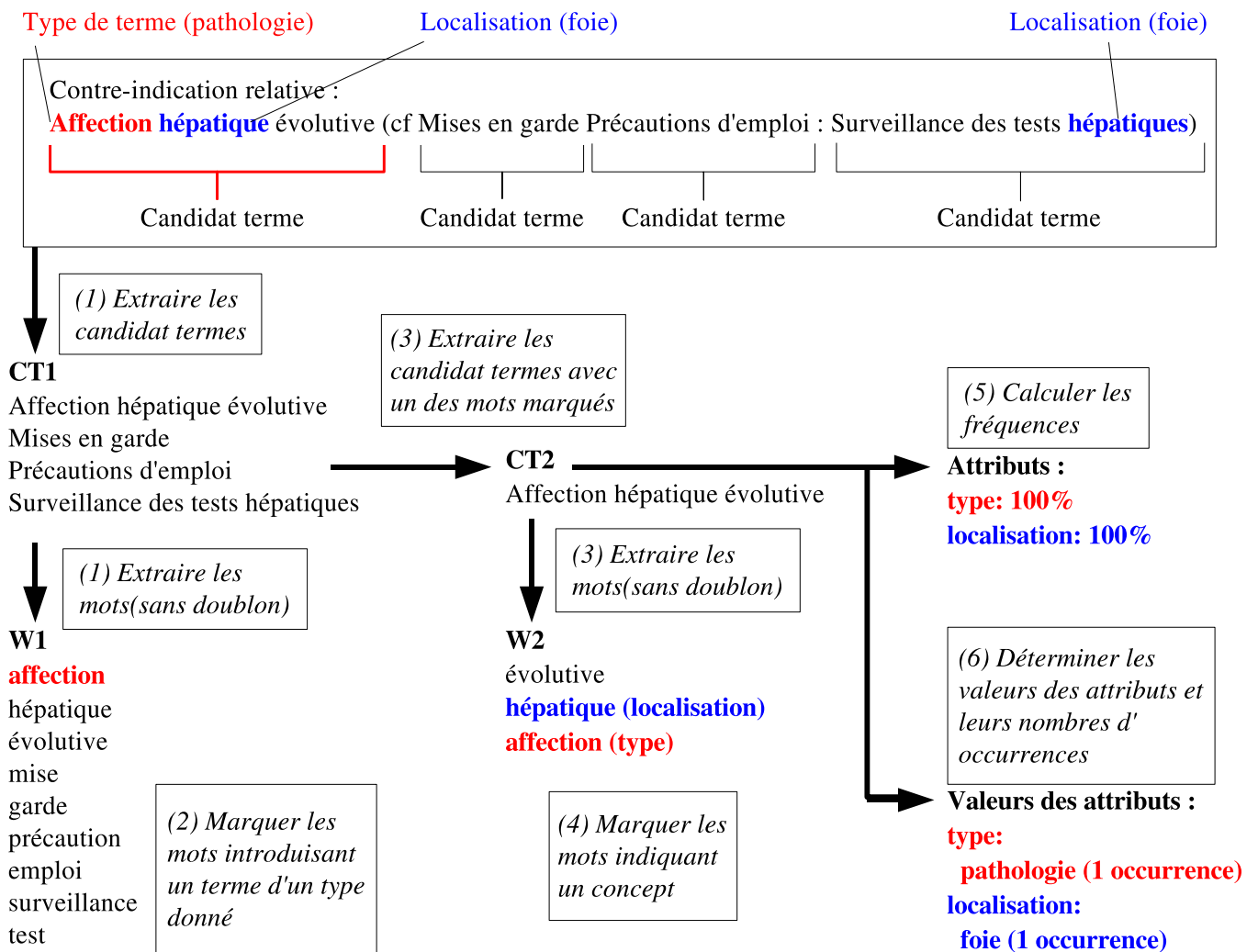


Fig. 4.1 – Méthode utilisée pour l'analyse sémantique des concepts médicaux. La méthode est ici appliquée sur les pathologies d'une seule phrase. Les numéros 1 à 5 correspondent aux étapes 1 à 5 dans le texte.

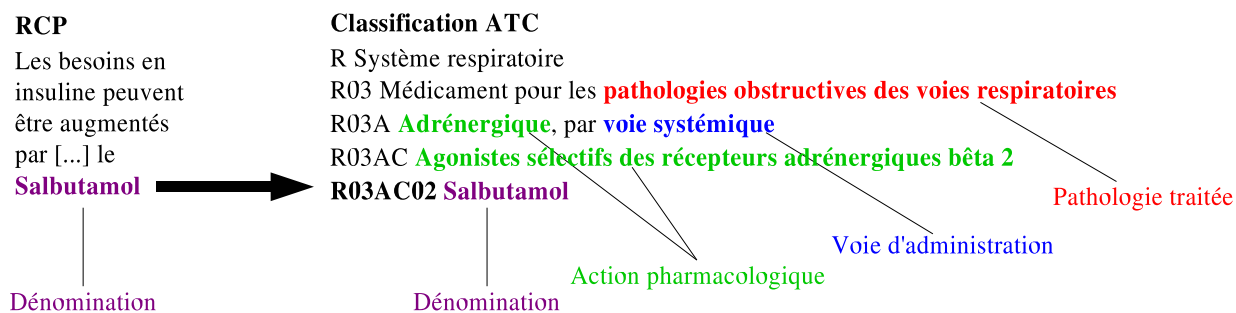


Fig. 4.2 – Exemple d’utilisation de la classification ATC pour enrichir les dénominations des médicaments. Ici, la dénomination “Salbutamol” est enrichie par des attributs apportant des informations sur l’activité pharmacologique, la voie d’administration et la pathologie traitée.

3. À partir de CT1, extraire la liste de tous les candidats termes comprenant au moins un mot marqué à l’étape précédente dans W1 ; nous l’appellerons CT2. Cette liste contient donc toutes les occurrences du type de concept qui nous intéresse (toutes les pathologies, par exemple). Ensuite, à partir de CT2, extraire la liste de tous les noms et adjectifs, sans doublon, que nous appellerons W2.
4. Marquer manuellement dans la liste W2 chaque mot avec la liste des attributs qu’il définit. Par exemple, “maladie” définit seulement l’attribut “type de concept” ; “cardiopathie” définit “type de concept” et “localisation” ; “hypertension” définit “type de concept”, “localisation” et “trouble” ; alors que “avocat” n’en définit aucun.
5. Pour chaque attribut identifié à l’étape 4, la fréquence de l’attribut est donnée par la proportion de candidat-termes dans la liste CT2 qui inclut au moins un mot marqué avec cet attribut dans la liste W2.
6. Pour chaque attribut identifié à l’étape 4, la liste des valeurs possibles est donnée par la liste des mots marqués avec cet attribut dans la liste W2, après avoir supprimé les doublons et regroupé manuellement les termes différents mais de même sens (par exemple pour l’attribut “localisation”, “coeur” et “cardiaque” ont la même signification). Le nombre d’occurrences d’un terme est donné par le nombre de fois où il apparaît dans la liste CT2.

Les étapes automatisées ont été réalisées avec Cordial et Lexter pour l’étape 1, et avec des scripts que j’ai écrit en langage Python pour les étapes 1, 3 et 5.

Dans le cas des dénominations, celles-ci n’indiquent pas directement les attributs du concept. Les dénominations ont donc été remplacées par les définitions correspondantes, par exemple le Salbutamol peut être considéré comme “un médicament agoniste sélectif bêta 2, par voie systémique, traitant les pathologies respiratoires obstructives”.

Ces dénominations concernent principalement les médicaments (DCI, nom de spécialité,...). Pour ceux-ci, la classification ATC et les relations *est-un* qu’elle contient ont été mises à profit. Les différents termes de la classification ATC ont été marqués avec les attributs qu’ils définissaient. Puis cette étude a été reliée à la précédente, en ajoutant pour chaque dénomination les attributs associés au terme ATC correspondant et à tous les termes ATC parents (voir figure 4.2).

4.2.3.3 Analyse de la structure des phrases

Les concepts médicaux ne sont pas utilisés de manière isolée mais font partie de phrases médicales ; par exemple une pathologie peut être utilisée au sein d’une contre-indication. En parcourant les RCP, il apparaît que ces phrases suivent des phrases-types qui varient relativement peu d’un RCP à l’autre. Cela n’est pas surprenant car, lors de la rédaction du RCP, l’utilisation de phrases-types est encouragée.

Une seconde analyse des RCP a été réalisée afin d’extraire les phrases-types les plus fréquentes. Voici l’algorithme utilisé (figure 4.3) :

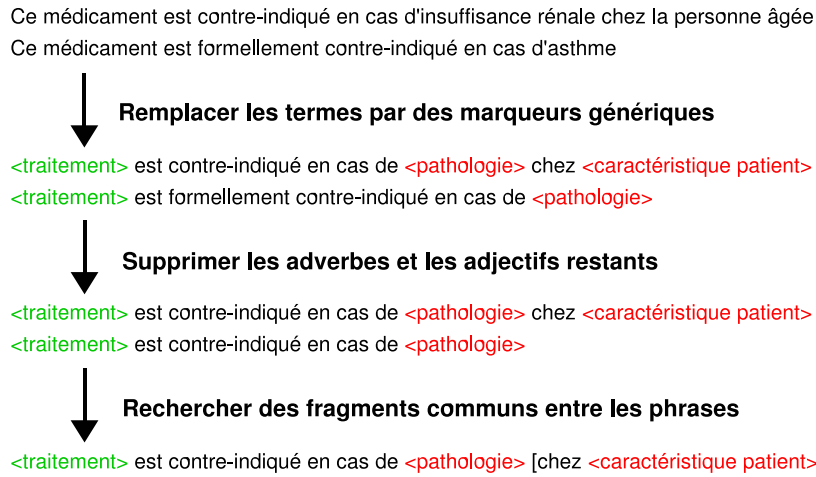


Fig. 4.3 – Méthode utilisée pour l'analyse de la structure des phrases.

La méthode est ici appliquée à deux phrases exprimant des contre-indications. Les crochets indiquent des morceaux de phrase optionnels, et les “<type de concept>” des variables qui peuvent être substituées par un concept du type indiqué.

1. Découper toutes les phrases, en se basant sur les marqueurs de ponctuation ‘. ; :’.
2. Lemmatiser chaque phrase et marquer chaque lemme avec sa fonction grammaticale, à l'aide de TreeTagger.
3. Remplacer chaque concept médical par un symbole (‘<état_du_patient>’, ‘<risque>’, ‘<médicament>’,...). Le remplacement a été réalisé en utilisant les listes de termes issues de l'analyse précédente.
4. Supprimer tous les adjectifs et adverbes restant. En effet, ceux-ci n'interviennent pas directement dans la structure de la phrase.
5. Chaque fragment (phrase ou morceau de phrase) A qui contient un autre fragment B, est découpé en une seconde occurrence de B et la partie restante de A ; les fragments comprenant moins de trois mots sont supprimés. Cette étape est effectuée en boucle, jusqu'à ce qu'il n'y ait plus de fragment découpable.
6. Les fragments sont triés en fonction de leur nombre d'occurrences.

Les phrases-types ont ensuite été reconstruites à la main à partir des fragments ainsi générés. Les phrases-types issus des différents corpus ont ensuite pu être combinées entre elles pour aboutir à une seule et unique phrase-type.

Enfin, l'utilisation des relations logiques ET et OU entre les concepts médicaux a été étudiée à partir d'un échantillon de RCP, afin de voir où ces relations sont employées et comment elles peuvent être imbriquées les unes dans les autres.

4.3 Résultats

4.3.1 Types de concepts

Dans les connaissances sur le médicament, les principaux types de concepts que nous avons identifiés sont les suivants :

- les états actuels du patient (pathologies, symptômes, états physiologiques, modes de vie...),
- les antécédents du patient (pathologies et symptômes passés),
- les risques (pathologies et symptômes futurs et possibles, dont effets indésirables),
- les médicaments et les propriétés des traitements médicamenteux (posologie,...),
- les surveillances médicales.

On retrouve parmi eux plusieurs notions importantes en médecine : la temporalité (passé : antécédents, présent : états actuels, futur : risques), l'incertitude (risques), les traitements (médicaments et surveillances),...

4.3.2 Attributs

Les concepts peuvent être décrits par des attributs, par exemple la localisation anatomico-fonctionnelle d'un état pathologique, ou la voie d'administration d'un médicament. Ces attributs peuvent ensuite prendre différentes valeurs, par exemple cardiaque ou rénale pour la localisation anatomico-fonctionnelle. La liste des différents attributs a été déterminée à partir de mes connaissances d'expert, puis étendue en s'inspirant des modèles existants, et à partir des éléments issus de l'analyse du corpus de RCP. Afin de déterminer la fréquence des attributs et leurs valeurs possibles, 6 analyses lexico-sémantiques ont été réalisées, chacune portant sur un corpus et un type de concept :

- sur le corpus contre-indication, les pathologies, symptômes et états physiologiques,
- sur le corpus effets indésirables, les risques,
- sur le corpus interactions médicamenteuses, les médicaments,
- sur le corpus précautions d'emploi-mises en garde, les surveillances médicales,
- sur le corpus précautions d'emploi-mises en garde, les caractéristiques du patient (âge, mode de vie,...),
- sur le corpus précautions d'emploi-mises en garde, les actions du médecin.

Les états actuels du patient ont été divisés en deux car les états physiopathologiques ne présentent pas les mêmes attributs que les caractéristiques du patient. Nous n'avons pas effectué d'analyse pour les antécédents, car ceux-ci diffèrent peu des pathologies actuelles et sont insuffisamment représentés (seulement 20-30 antécédents dans le corpus contre-indication). Le tableau 4.1 donne les différents attributs que nous avons identifiés, les valeurs qu'ils peuvent prendre avec leur nombre d'occurrences, la fréquence de chaque attribut lors de chacune des 5 analyses, ainsi que son importance médicale, sa facilité de représentation graphique, et si l'attribut a, ou non, été retenu pour la suite de notre travail.

Pour les états physiopathologiques (y compris les risques ou les antécédents d'états physiopathologiques), les attributs qui ont été retenus sont : l'état (pathologique ou non pathologique), la localisation anatomico-fonctionnelle et les troubles.

Les localisations anatomico-fonctionnelles sont représentées sur deux niveaux : un premier niveau correspondant au système (ex. : système respiratoire, système cardiaque), et un second niveau indiquant des structures anatomiques (ex. : bronches) ou des fonctions (ex. : rythme cardiaque) au sein de ce système.

Les troubles ont été divisés en deux catégories :

- les *troubles spécifiques* à une localisation donnée (ex. : les troubles du rythme sont spécifiques de la localisation cardiaque),
- les *troubles génériques*, qui peuvent se produire sur différentes localisations anatomico-fonctionnelles (ex. : l'insuffisance peut se produire au niveau cardiaque, rénal, thyroïdien), voire même sans localisation précise (ex. : infection sans plus de précision). Les troubles génériques peuvent indiquer une variation (augmentation ou diminution de fonctionnement d'une fonction physiologique, comme la fonction rénale), une étiologie (infection, cancer, allergie), un signe ou processus pathologique général / macroscopique (douleur, inflammation, oedème, hémorragie), un signe ou processus pathologique microscopique (atteinte vasculaire, atteinte nerveuse).

Pour les médicaments, les dénominations (DCI, noms de spécialité,...) ont été remplacées par les attributs correspondant en s'appuyant sur la classification ATC. Un code ATC a pu être trouvé pour 86% des dénominations, les autres correspondant à des familles chimiques très larges (acide, sel,...) ou sans équivalent ATC. Cette correspondance a été réalisée de façon automatique pour 56% des termes (qui correspondent aux DCI bien orthographiées), et de façon manuelle pour le reste (noms commerciaux, familles chimiques ou DCI comportant une faute d'orthographe).

attribut	valeurs possibles et nombre d'occurrences	patho	risque	méd	surv	carac	
<u>type de concept</u>	état actuel du patient (1100 pathologies + 1200 états non physiologiques et modes de vie, corpus (a)), antécédent, risque (7809, corpus (b)), médicament (8654, corpus (c)) ou surveillance (759, corpus (d))	100	100	100	100	100	RII
<u>dénomination (DCI, nom de spécialité,...)</u>	(nombreux noms, ex. : maladie de Parkinson, paracétamol,...)	2	2	71			II
<u>localisation anatomico-fonctionnelle</u>	sang 1442 (dont plaquette 645, sérum 220, globule blanc / immunité 208, globule rouge 103), psychologie / psychiatrie 1232 (dont humeur 336, sommeil 176, dépendance / sevrage 58), système digestif 1172 (dont estomac 515, intestin 388, bouche 108), système nerveux central 970, peau 751, cœur 513 (dont rythme cardiaque 260), respiration 355, système génital 263 (dont féminin 178, mâle 59), métabolisme 195 (dont glucidique 80), foie 339, rein 323, muscles 215, oeil / vision 159, grossesse 150, allaitement 133, os / articulation 102, oreille / audition 42	69	76	50	52		RII
<u>état (pathologique ou non)</u>	pathologique 18322, non pathologique 1200	100	100	100	100	100	RI
<u>trouble générique (pouvant se produire sur plusieurs localisations, par ex. une hypofonction / insuffisance peut se produire sur le coeur, le rein, le foie,...)</u>	hypofonction / diminution 1123, hyperfonction / augmentation 1097, douleur 433, arrêt 203, infection 2566 (dont virus 1069, bactérie 876, mycoses 220, parasites 135), tumeur 694, hypersensibilité / allergie 610, trouble circulatoire 588 (dont tension 404), thrombose / embolie 421, inflammation 396, trouble nerveux 347, oedème 113, hémorragie 71	60	46	50			RI
<u>trouble spécifique (spécifique d'une localisation donnée)</u>	(nombreuses valeurs selon les localisations (dont environ 50 pathologies majeures), par ex. les troubles du rythme pour le coeur)	14	34	15			RI
<u>effet (d'un médicament)</u>	(les effets sont opposés aux troubles : hypotenseur pour un anti-hypertenseur)			25			RI
<u>cible pharmacologique (d'un médicament)</u>	protéase 447, transcriptase reverse 356, adrénaline 223 (dont alpha 45, bêta 124), histamine 153 (dont H1 34, H2 38), choline 139, canaux calciques 129, monoamine oxydase 127, dopamine 108, HMG-COA 88, sérotonine 73,...			41			I
<u>activité (sur la cible pharmacologique)</u>	antagoniste 1834, agoniste 435			44			R
<u>sévérité (d'une pathologie)</u>	grave 375 (dont compliqué 67, malin 22), bénin 142 (dont isolé 61)	10	5				R
<u>fréquence (d'apparition d'un risque)</u>	rare 128, possible 78, observé 53, exceptionnel 39, fréquent 14		4				R
<u>substance spécifique (impliquée dans une pathologie, dosée lors d'une surveillance)</u>	créatine 72, glucose 49, transaminase 24, potassium 19, calcium 6, sodium 4, triglycéride 3, acétone 3, phosphore 2, magnésium 2,...	4	3		27		
<u>chronicité (d'une pathologie)</u>	aiguë 115, chronique 18	2	1				R
<u>réversibilité (d'un effet indésirable)</u>	réversible 78, irréversible 9		1				R
<u>voie d'administration (d'un médicament)</u>	systémique non précisé 1882, orale systémique 274, topique 161, injectable 71			26			R
<u>origine (du principe actif)</u>	plante 142 (dont millepertuis 57, digitale 51, ergot de seigle 19), virus 8			2			R
<u>méthode d'examen (d'une surveillance)</u>	examen biologique 414, clinique 94, fonctionnel 38, par imagerie 14				29		R
<u>examinateur (effectuant la surveillance)</u>	praticien 5, patient 3				1		R
<u>âge</u>	nourrisson 53, enfant 125, adulte 20, personne âgée 93					36	RI
<u>sexe</u>	femme 44, homme 3					5	RI
<u>mode de vie</u>	alcool 70, sport 39, alimentation 38, hygiène 11, tabac 8, conduite 3					22	RI
<u>action du médecin</u>	prescrire surveillance 692, prescrire médicament 389, posologie 364, prudence 345, arrêt du traitement 183, changer plan de prise, changer DCI						RI

Tab. 4.1 – Les attributs, leurs valeurs et leurs fréquences dans 5 types de concept. (états physiopathologiques actuels, risques, médicaments, surveillances et caractéristiques patient ; la case est vide si l'attribut ne s'applique pas à ce concept). La dernière colonne indique si l'attribut est représentable graphiquement (R), important (I) ou très important (II). Les attributs retenus sont soulignés.

Dans le langage naturel, il existe deux façons équivalentes d'indiquer l'action d'un médicament : par la pathologie qu'il traite (ex. anti-hypertenseur) ou par son effet (opposé à la pathologie traitée, ex. hypotenseur). L'analyse lexico-sémantique montre que, sur les médicaments du corpus interactions médicamenteuses, 65% des médicaments sont décrits par la pathologie traitée (trouble générique ou spécifique), alors que 25% sont décrits par leur effet (les 10% restant correspondent à des dénominations trop vagues pour définir un effet ou une pathologie traitée, par exemple des catégories chimiques comme "les acides"). C'est pourquoi nous avons choisi de représenter les médicaments par la pathologie qu'ils traitent, cette pathologie pouvant être représentée par les attributs déterminés précédemment pour les pathologies.

Pour les surveillances, les attributs retenus sont la localisation anatomico-fonctionnelle et la méthode d'examen.

Pour les caractéristiques du patient, seuls l'âge, le sexe et le mode de vie ont été retenus. Cela représente plus de 60% des caractéristiques patients; les autres caractéristiques (patient alité ou debout,...) étant très peu fréquentes.

Enfin, les **actions du médecin** peuvent être de deux types : soit la prescription d'un nouveau traitement ou d'une nouvelle surveillance, qui pourra être représentée comme le médicament ou la surveillance correspondant, soit la modification d'un traitement médicamenteux existant. La modification peut porter sur la posologie, le plan de prise, la forme pharmaceutique, la DCI ou l'arrêt du traitement [48, 47]. Il n'y a pas de fréquence associée à l'attribut "action du médecin", car cet attribut a fait l'objet d'une analyse pour lui tout seul.

En conclusion, les attributs suivants ont été retenus : **type de concept, localisation, état, trouble générique, trouble spécifique, voie d'administration, méthode de test, âge, sexe, mode de vie et modification du traitement.**

4.3.3 Relations

Parmi les concepts d'un même type, nous avons utilisé les classifications médicales existantes pour déterminer les relations *est-un*. Nous avons retenu la CIM10 pour les pathologies (état actuel, risque ou antécédent) et l'ATC pour les médicaments. Ces choix ont été motivés par la disponibilité de ces classifications et leur relative simplicité, puisque nous n'avons pas besoin d'un niveau de précision important.

Déterminer les relations *est-un* entre des concepts de types différents est plus compliqué. Dans les RCP et les GBP, comme dans la pratique courante, le langage naturel fait la confusion entre :

1. un état actuel (pathologique ou non) et le patient qui est dans cet état,
2. un risque et le patient qui présente ce risque,
3. un antécédent et le patient qui a cet antécédent,
4. un médicament et le patient qui prend ce médicament,
5. une surveillance et le patient qui est surveillé.

Par exemple (pour 4.) : "contre-indiqué avec l'amiodarone" signifie en fait "contre-indiqué chez les patients prenant de l'amiodarone". En clinique, les professionnels de santé sont plus intéressés par les relations entre leur patient et telle maladie ou tel médicament, que par les propriétés "académiques" de ceux-ci. Par conséquent, nous avons choisi de modéliser le patient souffrant de tel état, ayant tel risque ou tel antécédent, prenant tel médicament ou suivi par telle surveillance, et ensuite de représenter les états actuels, les risques, les antécédents, les médicaments ou les surveillances de la même manière. Deux relations *est-un* peuvent donc être déduites :

1. **un patient qui prend le médicament X traitant l'état Y est un patient dans l'état Y** (en général, Y est une pathologie ou un symptôme actuel),
2. **un patient qui est suivi par une surveillance X pour l'état Y est un patient dans l'état Y** (en général dans les connaissances sur le médicament, Y est un risque iatrogène).

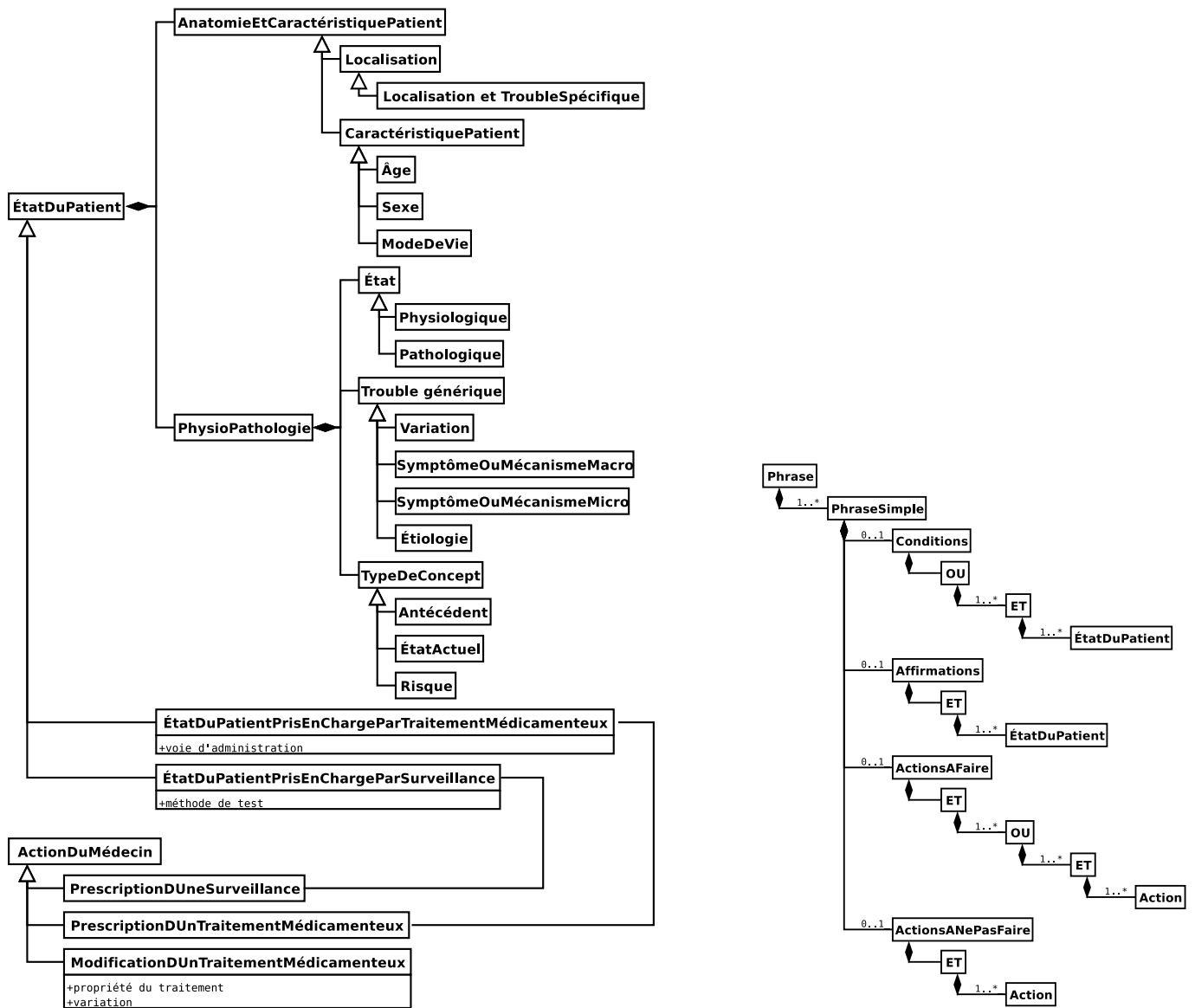


Fig. 4.4 – Modèle UML des états du patient (à gauche) et des phrases médicales (à droite).

Enfin, les troubles spécifiques ne peuvent pas être séparés des localisations correspondantes, par exemple il n'est pas possible de parler de "troubles du rythme" sans évoquer la notion de "cœur". Puisque le trouble spécifique implique la localisation, nous avons placé une relation d'héritage entre les deux ; c'est à dire que notre modèle représentera en fait "cœur" et "cœur atteint de troubles du rythme".

La figure 4.4 (à gauche) récapitule les différents concepts et leurs attributs dans le formalisme UML, ainsi que certaines relations *est-un*.

4.3.4 Phrases

En appliquant la méthodologie proposée précédemment (section 4.2.3.3) sur les 4 corpus, les phrases-types suivantes ont été obtenues (les crochets indiquent des parties optionnelles, et les éléments entre <> sont à substituer par un concept du type approprié) :

Corpus contre-indication

- ce médicament est contre-indiqué en cas de <caractéristique patient> [car il y a un risque de <risque>].
- [si <caractéristique patient>], ce médicament est contre-indiqué en cas de <état physiopathologique> [car il y a un risque de <risque>].

Corpus interactions médicamenteuses

- [si <caractéristique patient ou état physiopathologique>], ce médicament est contre-indiqué avec <médicament> [car il y a un risque de <risque>].

Corpus effets indésirables

- [si <caractéristique patient, état physiopathologique ou médicament>], il y a un risque de <risque>.

Corpus précautions d'emploi-mises en garde

- [si <caractéristique patient, état physiopathologique ou médicament>], il faut faire <action>.
- [si <caractéristique patient, état physiopathologique ou médicament>], il ne faut pas faire <action>.
- [si <caractéristique patient, état physiopathologique ou médicament>], il y a un risque de <risque> [, pour éviter ce risque il faut faire <action>].

Voici un exemple pour chacune des phrases-types :

contre-indication : contre-indiqué chez la personne âgée.

contre-indication : chez les personnes âgées, contre-indiqué en cas d'insuffisance rénale.

interactions médicamenteuses : chez le sujet âgé, contre-indiqué avec les héparines.

effets indésirables : risque de troubles hépatiques.

précautions d'emploi : en cas de vomissements, il est recommandé de prescrire des anti-émétiques.

précautions d'emploi : chez les angineux, ne jamais interrompre brutalement le traitement.

précautions d'emploi : avec le diltiazem injectable, risque de bradycardie ; si l'association ne peut être évitée, surveillance de l'ECG.

Dans les conditions, la condition "prendre le médicament du RCP" est implicite et ne figure pas ici ; celle-ci n'est que très rarement explicitée dans le texte des RCP.

Les différentes phrases-types présentent un important degré de similitude, et partagent des fragments communs, exprimant des conditions (si...), des affirmations (il y a un risque de...), des conseils (il faut faire...) ou des actions à ne pas faire (il ne faut pas faire..., ce médicament est contre-indiqué = il ne faut pas prescrire ce médicament). L'ensemble des phrases-types précédentes peuvent donc être regroupées en une seule :

[si <caractéristique patient, état physiopathologique ou médicament>] [il y a un risque de <risque>] [il faut faire <action>] [il ne faut pas faire <action>].

L'analyse manuelle des relations logiques ET et OU a donné les résultats suivants :

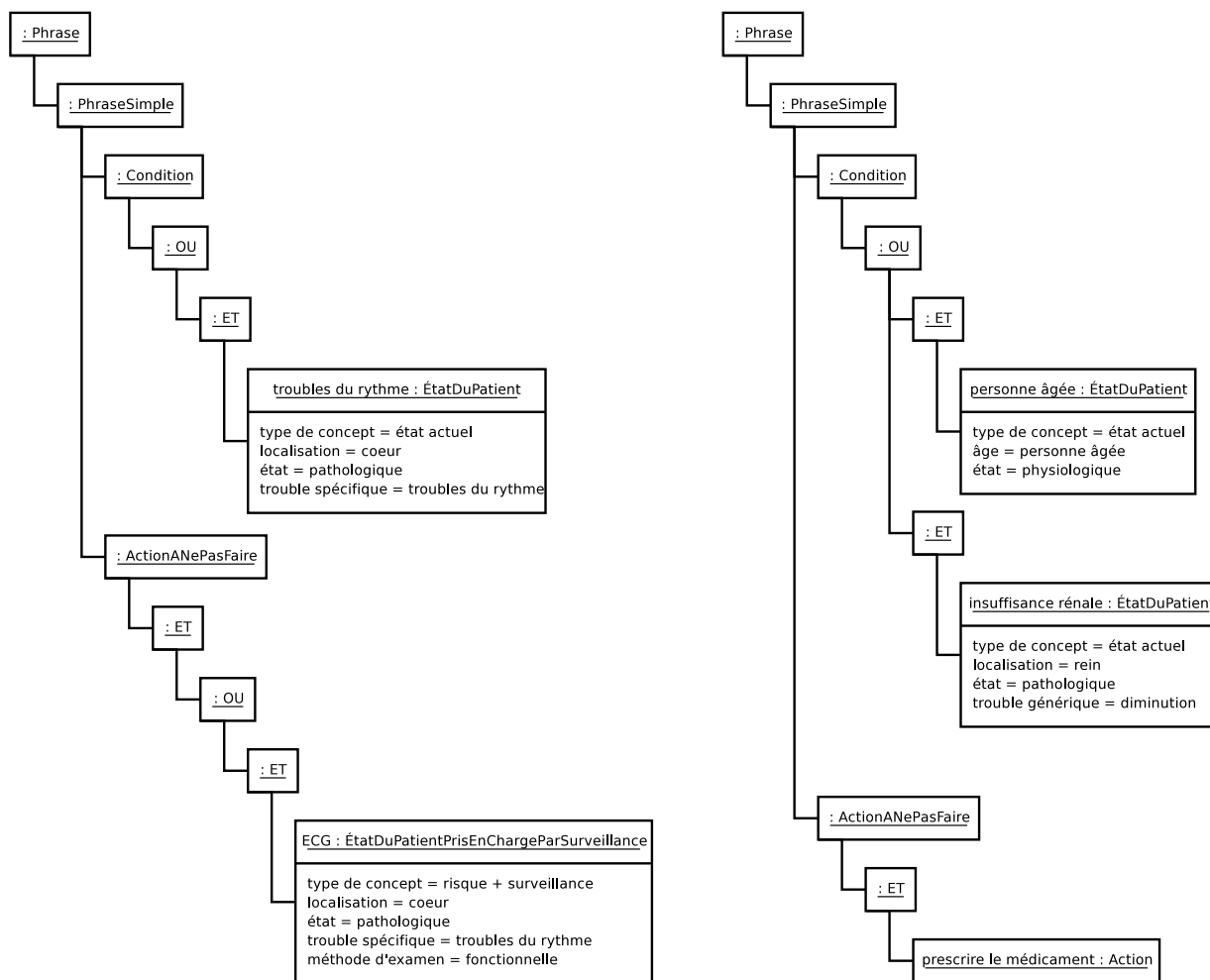


Fig. 4.5 – Exemples d’utilisation de notre modèle.

Les deux exemples correspondent aux phrases “en cas de troubles du rythme, il est conseillé d’effectuer un ECG” et “(ce médicament est) contre-indiqué chez la personne âgée *et* en cas d’insuffisance rénale” (notons que ce “*et*” a en fait la valeur d’un OU logique), et sont représentés en UML.

Dans les conditions l’organisation de ces relations correspond au motif : “si (a ET b) OU (c ET d)” (ex. : “Il est recommandé d’éviter toute déshydratation préalable à l’examen et de maintenir une diurèse abondante chez les insuffisants rénaux, ou chez les sujets âgés athéromateux”, RCP du Iopamiron 200 solution injectable).

Dans les affirmations et les actions à ne pas faire seules des relations ET ont été rencontrées.

Dans les conseils les relations ET et OU sont imbriquées soit dans un sens, soit dans l’autre. En conséquence, nous avons retenu le motif plus général : “il faut faire ((a ET b) OU (c ET d)) ET ((e ET f) OU (g ET h))”.

Ensuite, des phrases plus complexes peuvent être construites en assemblant plusieurs phrases simples. La figure 4.4 récapitule la structure des phrases (à droite), et la figure 4.5 montre deux exemples d’application de ce modèle.

4.4 Discussion et conclusion

4.4.1 À propos de la méthode

La méthode que nous avons utilisée repose à la fois sur des connaissances issues d’experts ou de la littérature, et sur l’analyse des textes par des techniques de TAL. Dans la littérature, d’autres auteurs ont associé ces deux aspects pour concevoir des modèles de connaissances médicales [43, 14].

Les classifications médicales (voir section 3.5) proposent une hiérarchisation des termes médicaux, en particulier celles ayant une structure arborescente (CIM10, ATC). Cependant, cette hiérarchisation n'est pas toujours homogène : à un même niveau de profondeur dans la classification, nous trouvons des termes ayant des importances médicales très diverses (par exemple au niveau 3 de la CIM10 "I10 hypertension essentielle (primitive)" et "I05 maladies rhumatismales de la valvule mitrale"). C'est pourquoi nous n'avons pas utilisé ces classifications pour déterminer l'importance des concepts, en décidant par exemple de représenter toutes les pathologies des 2 ou 3 premiers niveaux de la CIM10.

L'utilisation de techniques de TAL a permis d'obtenir des informations chiffrées sur la nature des connaissances sur le médicament, et la manière dont elles sont utilisées dans les RCP. Ces résultats ont été décisifs à plusieurs reprises, notamment pour le choix du mode de représentation des médicaments (pathologie traitée ou effet ?). Par contre, ces analyses se sont révélées très coûteuses en temps, à cause des étapes manuelles 2. et 4. de la section 4.2.3.2. Nous aurions sans doute pu gagner du temps en restreignant ces analyses aux seuls points délicats, plutôt que de les réaliser pour l'ensemble des attributs.

Le corpus choisi, les RCP français correspondant aux médicaments essentiels de l'OMS, favorise certaines classes thérapeutiques, en particulier les traitements contre le Sida. Cependant, ce fait étant connu, nous l'avons pris en compte lors de l'interprétation des résultats. Le choix d'un corpus de RCP en langue française a pu avoir une influence sur les résultats obtenus ; nous discuterons de l'indépendance de VCM par rapport aux langues et aux cultures à la section 8.3.3.

En ce qui concerne la méthode de TAL proprement dite, la méthode que nous avons proposée présente deux avantages. Tout d'abord, elle ne prend en compte que les attributs vraiment utilisés dans l'expression d'un concept, et non en-dehors (par exemple sur la figure 4.1, le mot "hépatique" apparaît deux fois, mais seule la première occurrence concerne un état physiopathologique ; la seconde occurrence ne sera donc pas prise en compte). Ensuite, les étapes manuelles portent uniquement sur des listes de mots, et non de syntagmes nominaux, ce qui permet de gagner du temps (il y a en pratique moins de mots différents que de syntagmes nominaux différents).

Nous n'avons pas évalué la complétude du modèle obtenu, c'est-à-dire la proportion des connaissances sur le médicament qu'il est capable de représenter. En effet, la complétude du modèle est volontairement faible, l'objectif étant de disposer d'un modèle simplifié. Plutôt que la complétude, il serait intéressant d'évaluer la pertinence des connaissances retenues par le modèle, ce qui n'est pas facile à déterminer car très subjectif. Nous avons donc fait le choix d'évaluer le modèle au travers du langage graphique VCM, lors de l'évaluation du langage auprès des médecins.

4.4.2 À propos des résultats

Le modèle que nous avons conçu présente deux particularités qui le distinguent de ceux de la littérature. Premièrement, **le modèle est beaucoup plus simple et moins détaillé que les modèles existants** ; cela le rend utilisable pour la construction d'un langage graphique, qui est nécessairement moins précis que le texte.

La seconde particularité réside dans la confusion qui est faite entre traitements (médicamenteux ou par surveillance) et patients traités par ces traitements, et **le choix délibéré de représenter le patient traité plutôt que le traitement**, comme cela est fait le plus souvent (par exemple dans le modèle des interactions médicamenteuses de la section 3.3, l'interaction se fait avec un médicament et non un patient prenant le médicament). Les modèles de GBP à base de scénarios proposent une approche similaire, en ramenant tous les concepts (pathologies, traitements,...) au patient. Nous pensons que ce choix est intéressant, car il nous semble être plus proche du mode de pensée des professionnels de santé.

Le modèle des phrases médicales que nous proposons est intéressant car ce modèle est capable de représenter les phrases issues des différentes sections cliniques du RCP (contre-indications, interactions médicamenteuses, précautions d'emploi-mises en garde, effets indésirables), alors que la plupart des modèles existants proposent un modèle différent pour chaque section. Ce point est important car, dans un langage graphique, il serait fastidieux d'avoir plusieurs grammaires différentes pour les

phrases, selon la section du RCP.

4.4.3 Conclusion

En nous appuyant sur des connaissances d'expert du domaine, sur la littérature et sur une analyse d'un corpus de RCP par des méthodes de TAL, nous avons construit un modèle des connaissances sur le médicament. Par sa simplicité, ce modèle est bien adapté à la conception d'un langage graphique.

Nous allons maintenant voir comment construire le langage graphique VCM, en s'appuyant sur notre modèle des connaissances sur le médicament et sur les aspects de sciences cognitives que nous avons étudiés au chapitre 2.

Chapitre 5

Conception du langage graphique de visualisation des connaissances sur le médicament

5.1 Introduction et objectif

Une fois que la nature et la structure des connaissances sur le médicament ont été déterminées, nous pouvons construire le langage VCM. Comme tout langage, celui-ci se compose d'une grammaire et d'un lexique. La construction de la grammaire devra prendre en compte les considérations issues de la sémiotique et des différents modèles de la vision humaine (voir section 2.1), afin de traduire les attributs, les concepts et les phrases médicales dans un langage graphique qui soit à la fois rapide à lire et facile à apprendre.

La conception du langage VCM s'est faite de manière itérative, en produisant un grand nombre de variantes et de versions intermédiaires du langage, et en affinant progressivement la méthodologie utilisée. Seule la méthodologie la plus avancée et le langage VCM stabilisé sont présentés ici ; l'évolution du langage est résumée en annexe D.

5.2 Matériel et méthodes

Le langage médical textuel, comme les modèles des connaissances médicales, contient une part d'information implicite : par exemple, lorsque l'on parle d'angor ou de troubles du rythme, il s'agit implicitement d'un problème cardiaque. Or c'est cette part d'information implicite qui rend l'apprentissage plus difficile, car il faut apprendre par coeur par exemple que l'angor touche le coeur. Elle complique aussi les recherches au sein d'un document, par exemple si un médecin recherche tout ce qui se rapporte à des problèmes cardiaques dans un texte, il ne peut pas limiter sa recherche au mot "coeur" et à la racine "cardi-", sans quoi il passera à côté de l'angor et des troubles du rythme. Lors d'une recherche, la présence d'information implicite oblige à lire et à interpréter chaque mot, afin d'en comprendre le sens et de déterminer si oui ou non il se rapporte au coeur. Or cette interprétation est coûteuse en temps, car elle se déroule au second niveau d'interprétation (SNC, voir figure 2.6).

Nous ferons donc l'hypothèse suivante :

Hypothèse 1 :

Un langage graphique est plus facile à apprendre et plus rapide à lire, s'il représente de manière explicite des informations ou des connaissances qui sont implicites dans le langage textuel ou dans les modèles de connaissances.

En effet, si une information est représentée explicitement par le langage, d'une part le médecin n'aura pas besoin de l'apprendre par coeur, et d'autre part il n'aura pas toujours besoin de faire appel à sa mémoire lors de l'interprétation du langage (nous avons vu à la section 2.1.6 que le premier niveau d'interprétation, le plus rapide, n'a qu'un accès très limité à la mémoire).

	universalité	apprentissage		vitesse de lecture
		compréhension	mémorisation	
symbole arbitraire	-	-	+	++
convention	-	++	++	++
analogie	+	+	++	+

Tab. 5.1 – Avantages et inconvénients des analogies, conventions et symboles arbitraires. “-” signifie que la propriété est absente et “+” qu’elle est présente ; plus le nombre de “+” est grand, plus celle-ci est forte.

Par ailleurs, il est possible de faciliter l’apprentissage du langage graphique en utilisant un lexique et une grammaire simple :

Hypothèse 2 :

Le lexique d’un langage graphique est plus facile à apprendre, si le lexique repose sur l’analogie ou sur des conventions déjà connues par l’utilisateur.

Hypothèse 3 :

La grammaire d’un langage graphique est plus facile à apprendre et plus rapide à lire, si la grammaire décompose les connaissances de la même manière que la perception visuelle décompose les phrases graphiques du langage.

Tout d’abord, nous étudierons l’utilisation de l’analogie et des conventions connues. Nous mettrons ensuite notre hypothèse en pratique pour représenter graphiquement de manière explicite les attributs des concepts, à l’aide d’une grammaire combinatoire, ainsi que les relations *est-un* entre les concepts. Puis nous verrons comment combiner les différents signes graphiques tout en faisant ressortir la grammaire sous-jacente. Enfin, nous verrons comment concevoir les pictogrammes qui composent le lexique du langage. Dans chacune de ces parties, nous essaierons de donner des règles pour la conception d’un langage graphique, que nous appliquerons ensuite à la création de VCM. Pour terminer, nous proposerons un formalisme pour la grammaire d’un langage graphique.

5.2.1 Analogie, conventions et symboles arbitraires

Nous avons vu lors de l’état de l’art (section 2.1.1) deux méthodes de représentation possibles : l’analogie utilise un signe iconique (signe ressemblant à l’objet représenté) ou un index (signe ressemblant à un objet relié à l’objet représenté), tandis que le symbole est un signe arbitraire sans rapport avec l’objet représenté et qui doit faire l’objet d’un apprentissage. De plus, les symboles peuvent faire l’objet de conventions bien connues au sein d’un domaine, ou non. À titre d’exemple, le sexe féminin peut être représenté :

de manière analogique par une silhouette avec une jupe,

à l’aide d’une convention par le symbole ♀,

à l’aide d’un symbole arbitraire par un triangle.

Ces trois méthodes de représentation ont chacune des avantages et des inconvénients, récapitulés dans le tableau 5.1. L’universalité indique la capacité du signe à être reconnu de manière “universelle” par des personnes “naïves” rencontrant ce signe pour la première fois et ne connaissant aucune convention, par exemple des personnes d’une autre culture. Seule l’analogie permet de concevoir des signes “universels”.

La compréhension est liée à l’universalité, mais suppose que la personne connaît les conventions du domaine : il s’agit de la capacité du signe à être reconnu lors de sa première rencontre par une personne du domaine. La mémorisation indique la facilité à se rappeler la signification du signe pour les lectures ultérieures. Compréhension et mémorisation sont deux composantes de l’apprentissage du signe : pour apprendre le signe, il faut d’abord le comprendre puis le mémoriser. L’apprentissage des symboles arbitraires est difficile, puisque, étant arbitraire, il n’est pas possible de comprendre leur signification. Au contraire, l’apprentissage des signes analogiques est facilité par leur aspect universel :

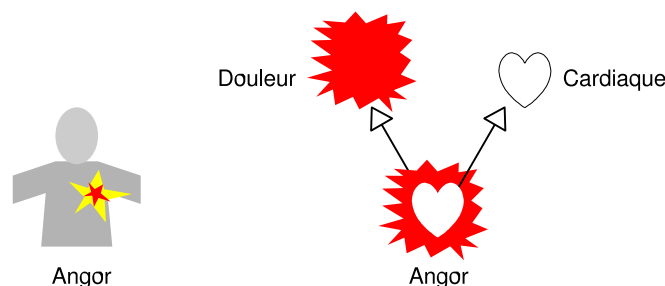


Fig. 5.1 – Exemples de représentation de l’angor (ou angine de poitrine) sans ou avec une grammaire combinatoire.

À gauche, l’angor est représenté par un signe dessiné spécialement pour cette pathologie ; à droite l’angor est décomposé en deux attributs (trouble générique et localisation).

il est possible de comprendre le sens du signe, et cette compréhension facilite la mémorisation. Enfin, les conventions, si elles sont déjà connues, ne nécessitent aucun apprentissage nouveau.

La vitesse de lecture est la plus grande pour les signes arbitraires, car, étant choisis de manière arbitraire, il n’y a que très peu de contraintes sur le choix du signe, et les symboles les plus simples, et donc les plus faciles à lire, pourront être choisis. Au contraire les signes analogiques doivent ressembler à ce qu’ils représentent, ce qui rend difficile l’usage de signes simples. Bien entendu, cela peut varier d’un signe à l’autre en fonction de l’objet à représenter, et dans certains cas, un signe analogique peut être aussi rapide à lire qu’un symbole.

Pour la conception du langage VCM, l’universalité nous intéresse peu puisque nous nous adressons à des professionnels de santé qui disposent donc d’un bagage médical, de plus l’aspect “multi-culturel” n’est pas l’objectif premier de VCM (voir section 1.2 et 8.3.3). Seule la facilité d’apprentissage et la vitesse de lecture nous intéressent. C’est pourquoi nous utiliserons au maximum l’analogie et les conventions déjà connues des professionnels de santé. Le choix entre analogie et conventions existantes, lorsque les deux sont possibles, est délicat et doit être traité au cas par cas. Dans certains cas, l’analogie conduirait à des signes trop complexes ; dans d’autres, les conventions existantes peuvent être difficiles à inclure dans un langage plus vaste.

Règle 1 :

Un langage graphique destiné aux professionnels de santé doit utiliser au maximum l’analogie et les conventions existantes dans le milieu médical.

En revanche, les symboles arbitraires sont difficiles à apprendre et ne doivent pas être utilisés, sauf de manière très restreinte, et pour des signes utilisés fréquemment. En effet, si ces symboles sont utilisés régulièrement, cela facilitera leur mémorisation.

Règle 2 :

Les symboles arbitraires ne doivent être utilisés que lorsqu’il n’est pas possible de faire autrement, ou bien pour un très petit nombre de symboles différents, utilisés fréquemment par le langage.

5.2.2 Grammaire combinatoire

Lors de la modélisation des connaissances au chapitre 4, nous avons décrit les différents concepts médicaux par plusieurs attributs, par exemple l’angor pourra être décrit par les attributs *trouble générique* et *localisation anatomico-fonctionnelle*, prenant respectivement les valeurs *douleurs* et *cardiaques*. Pour représenter graphiquement ces concepts, une première possibilité consiste à dessiner un signe spécifique pour chaque concept, par exemple un signe pour l’angor, un autre pour troubles du rythme,... Dans ce cas, il suffit de reconnaître ce signe pour le comprendre, et il n’y a alors pas de grammaire à proprement parler : le langage se réduit à un lexique.

Cependant il est aussi possible de représenter de manière explicite cette décomposition en attributs : au lieu de dessiner un signe pour chaque concept, nous pouvons représenter chaque concept par

une combinaison de plusieurs signes correspondant chacun à un attribut. Par exemple, l'angor pourra être représenté par la combinaison d'un signe représentant la douleur et d'un signe représentant le coeur (figure 5.1). Dans ce cas, des règles de grammaire sont nécessaires pour permettre d'identifier le rôle joué par chaque signe : quel signe indique le trouble ? et la localisation ? Nous parlerons alors de **grammaire combinatoire**.

Cette grammaire devra être apprise par l'utilisateur du langage graphique. En revanche, elle permet de rendre explicites les attributs qui définissent le concept. Rendre explicite cette information qui était implicite facilite l'apprentissage du langage graphique de deux manières. Tout d'abord, les icônes sont construites comme des "listes des attributs" qui définissent les concepts représentés. L'icône est une *définition* graphique du concept : par exemple "patient souffrant d'angor" sera représenté par une icône qui ne signifiera pas "patient souffrant d'angor" mais "patient souffrant actuellement d'une douleur localisée au coeur". Puisque chaque icône est une définition, le langage graphique devient alors une sorte "d'ontologie graphique".

Ensuite, cela permet de diminuer le nombre de signes à apprendre, lorsque le nombre de signes utilisés pour construire les définitions avec la grammaire combinatoire est inférieur au nombre total de concepts. Par exemple pour représenter les troubles du rythme, nous pourrions ré-utiliser le signe représentant la localisation cardiaque utilisé pour l'angor. Dans certains cas cependant, le nombre de signes total peut être supérieur au nombre de concepts, ce qui rend la grammaire combinatoire inutile.

Règle 3 :

Il faut utiliser une grammaire combinatoire lorsque les concepts à représenter sont décrits par plusieurs attributs, et lorsque la grammaire permet de diminuer notablement le nombre d'éléments à mémoriser pour l'apprentissage du langage graphique, c'est-à-dire lorsque le nombre de valeurs que peuvent prendre les attributs est inférieur au nombre de concepts.

En pratique, les signes représentant les attributs se limiteront à une seule variable rétinienne (couleur, forme, taille,...), afin d'être le plus simples possibles (voir section 2.1.3 pour la définition et la liste des variables rétinienne). Le choix de la variable qui représentera tel ou tel attribut doit prendre en compte les analogies possibles et les conventions connues (voir section précédente), ainsi que les propriétés des variables rétiniennes, en particulier le nombre de valeurs qu'elles peuvent prendre (ce que J. Bertin appelle la "longueur") :

Règle 4 :

Pour représenter un attribut, il faut choisir une variable rétinienne capable de représenter un nombre de valeurs au moins égal au nombre de valeurs que peut prendre l'attribut. La même variable rétinienne peut être utilisée pour plusieurs attributs différents.

En ce qui concerne les connaissances médicales, l'utilisation de grammaire combinatoire apparaît intéressante car le nombre de concepts à représenter est important. En particulier, lors de l'analyse des attributs (section 4.3.2), nous avons décrit les pathologies par leur localisation anatomico-fonctionnelle et les troubles qu'elles provoquent, avec un nombre élevé de valeurs possibles (environ 35 localisations et plus de 50 troubles). Les troubles ont été séparés en deux catégories, selon qu'ils sont génériques ou spécifiques d'une localisation. Les troubles spécifiques ne peuvent être appliqués qu'à une seule localisation : dans ce cas, l'utilisation d'une grammaire combinatoire n'apporte rien, et il est plus simple de créer un signe propre pour chaque couple (localisation, trouble spécifique).

En revanche, pour les troubles génériques, plusieurs troubles différents peuvent s'appliquer à la même localisation, et un même trouble peut s'appliquer à plusieurs localisations différentes. Dans ce cas, l'utilisation d'une grammaire combinatoire est pertinente. Avec 35 localisations et 20 troubles génériques, il y a un total de 700 combinaisons possibles ; cependant toutes ces combinaisons ne sont pas toutes intéressantes. Nous pouvons distinguer 3 catégories :

Les combinaisons utiles ont un sens médical et correspondent à des pathologies existantes. Par exemple : douleur au niveau du coeur (angor),...

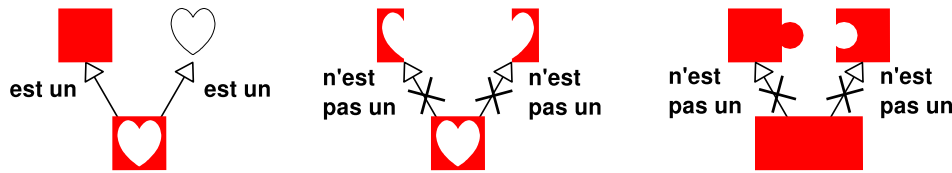


Fig. 5.2 – Exemples d'utilisation de la théorie de la *Gestalt* pour déterminer les relations *est-un* graphiques. À gauche, le carré et le coeur sont des formes simples, connues et bien délimitées par la couleur, contrairement à la décomposition du milieu. De la même manière, un rectangle ne peut pas être décomposé en “pièces de puzzle” car celles-ci ne sont pas visibles dans le rectangle.

Les combinaisons cohérentes mais inutiles pourraient avoir un sens, mais n'existent pas en pratique. Par exemple : hémorragie des organes sexuels masculins, atteinte nerveuse au niveau du foie,...

Les combinaisons incohérentes n'ont pas de sens médical. Par exemple : cancer du sommeil, infection de l'humeur, douleur localisée aux globules rouges,...

L'existence de combinaisons incohérentes n'est pas un problème en soi ; la langue française a elle-même beaucoup de combinaisons incohérentes, par exemple “oireghoreg” est une combinaison incohérente de lettres. Cependant, l'intérêt de la combinatoire est d'autant plus grand que le rapport nombre de combinaisons utiles sur nombre de signes de base est grand. Afin d'évaluer le nombre de combinaisons utiles, et de vérifier que celui-ci est bien supérieur au nombre de signes utilisés pour représenter les localisations et les troubles génériques, j'ai passé en revue les 700 combinaisons possibles et je les ai classées manuellement selon les 3 catégories.

La combinatoire peut aussi être utilisée pour les autres attributs du langage : par exemple l'attribut “type de concept” sera représenté par un signe indiquant s'il s'agit de l'état actuel du patient, d'un antécédent, d'un risque, d'un état actuel traité par un médicament ou d'un risque suivi par une surveillance. Il en est de même pour les attributs état pathologique, voie d'administration, méthode d'examen, âge, sexe, mode de vie et modification du traitement.

5.2.3 Représentation graphique des relations *est-un*

Nous avons identifié un grand nombre de relations *est-un* dans les connaissances médicales, et celles-ci ont été incluses dans notre modèle. D'autres relations *est-un* apparaissent lors de l'utilisation de la grammaire combinatoire : si nous reprenons l'exemple de l'angor, l'angor *est une* pathologie cardiaque et *est une* douleur.

Les relations *est-un* font partie des informations qui sont souvent implicites dans le langage médical ; par exemple l'angine de poitrine *est une* pathologie cardiaque, mais cela est implicite. **Il est important d'explicitier ces relations et de les représenter graphiquement pour faciliter l'apprentissage et l'utilisation du langage graphique.** En effet, lorsqu'un médecin s'intéresse à un concept, par exemple aux pathologies cardiaques, il est aussi intéressé par les concepts fils, par exemple les troubles du rythme. S'il recherche l'icône VCM des pathologies cardiaques à l'écran ou sur une feuille de papier, il doit donc pouvoir trouver l'icône des troubles du rythme. Cela n'est possible que si le signe qui représente les troubles du rythme *est un* signe qui représente les pathologies cardiaques, c'est à dire que lorsque le médecin regarde le signe des troubles du rythme, il doit y voir aussi le signe des pathologies cardiaques.

Les lois de ségrégation issues de la théorie de la *Gestalt* (voir section 2.1.2) permettent de déterminer si un signe est visible dans un autre. Si (et seulement si) le signe B est visible dans le signe A, alors nous dirons que le signe A *est un* signe B (relation *est-un* graphique, voir figure 5.2 et la présentation effectuée lors d'un atelier [88]). En particulier, il ne suffit pas “d'ajouter” ou de superposer deux signes A et A', donnant naissance au signe A+A', pour obtenir les relations A+A' *est un* A, et A+A' *est un* A'. En effet, rien ne garantit qu'après la superposition, les signes A et

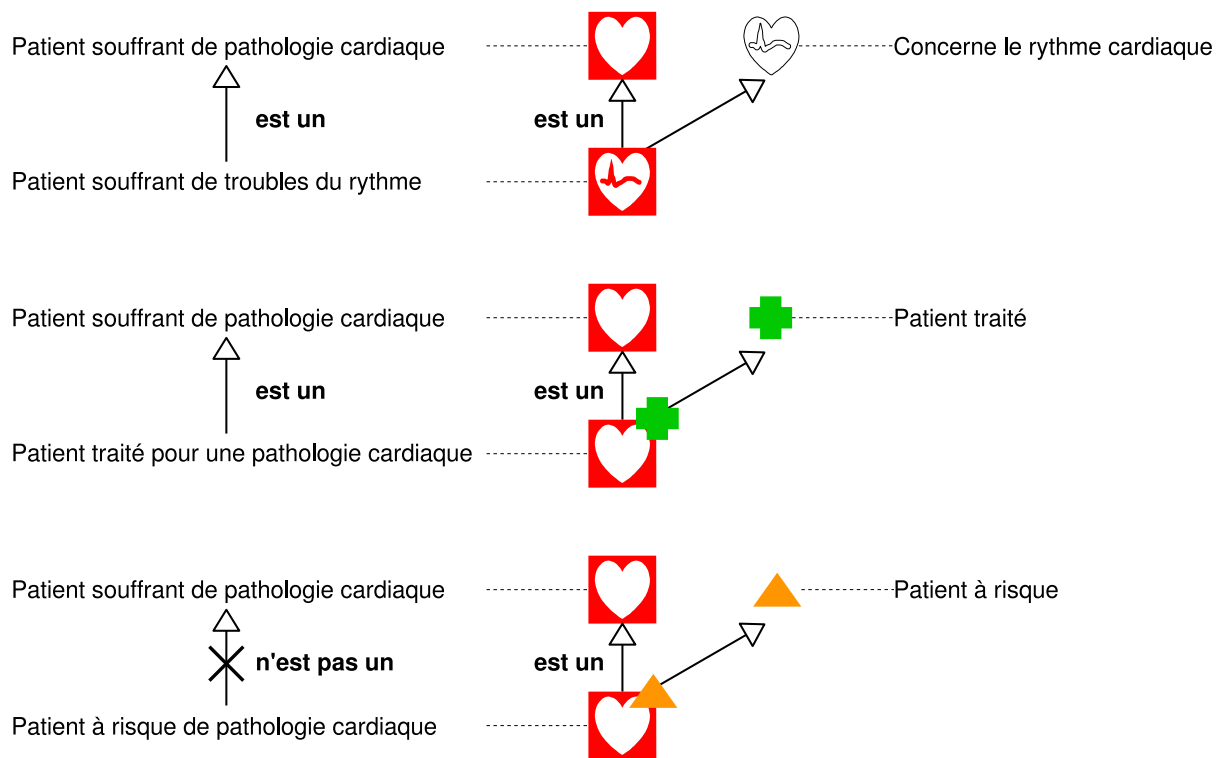


Fig. 5.3 – Exemples de relations *est-un* graphiques.

Le dernier exemple montre ce qu'il ne faut pas faire.

A' soient toujours visibles dans le signe A+A' (voir exemple des pièces de puzzle figure 5.2, et le principe fondamental de la théorie de la *Gestalt* : un tout n'est pas la somme de ses parties).

Nous pouvons donc établir la règle suivante :

Règle 5 :





Si (et seulement si) le concept A *est un* concept B, alors le signe qui représente le concept B doit être visible dans le signe qui représente le concept A. Les lois de ségrégation doivent être utilisées pour déterminer ce qui est visible dans un signe, et ce qui ne l'est pas.

Par exemple (figure 5.3), nous pouvons représenter les troubles du rythme par le signe de la pathologie cardiaque, et le signe du rythme cardiaque (notons que le coeur est présent dans ces deux signes, ce qui n'est pas un problème). De même, nous pouvons représenter le patient traité pour une pathologie cardiaque par le signe de la pathologie cardiaque et le signe du patient traité.

En revanche, nous ne pouvons pas représenter le patient à risque cardiaque par le signe du patient souffrant de pathologie cardiaque et le signe du patient à risque. En effet, ce signe *est un* signe du patient souffrant de pathologie cardiaque, or le patient à risque cardiaque n'est pas un patient souffrant de pathologie cardiaque ! Si nous utilisions cette manière de représenter le risque, lorsque le médecin rechercherait des informations concernant un patient atteint de pathologie cardiaque, il trouverait aussi les informations concernant les patients à risques cardiaques, alors qu'il ne les recherche pas. Nous devons donc représenter le risque d'une autre manière, par exemple en remplaçant la couleur rouge par une autre couleur, de sorte à ne plus avoir sur l'icône du risque le carré rouge avec un coeur à l'intérieur (voir résultats).

5.2.4 Méthodes de combinaison de signes

Lorsque l'on utilise une grammaire combinatoire ou que l'on représente graphiquement des relations *est-un*, il est nécessaire de combiner ensemble plusieurs signes, par exemple ceux représentant les différents attributs dans la grammaire combinatoire. Nous allons voir tout d'abord les différentes méthodes de combinaison de signe, et ensuite nous verrons comment choisir entre ces méthodes.

méthodes de combinaison de signe		séparabilité	contraintes sur les signes	Exemples
juxtaposition	sans recouvrement	++++	(aucune)	
	avec recouvrement	+++	(aucune)	
inclusion		++	la forme du signe placé à l'extérieur doit être suffisamment simple (forme extérieure)	
combinaison des variables rétinienne		+	les deux signes doivent utiliser des variables rétiniennes différentes	

Tab. 5.2 – Propriétés des différentes méthodes de combinaison de signes.

La séparabilité indique la propension à faciliter ou non la distinction entre les deux signes ; plus le nombre de “+” est grand, plus il est facile de distinguer les deux signes après combinaison. Pour les 3 premières lignes, les exemples combinent une flèche vers le bas et un rein dans le but de signifier “insuffisance rénale” ; pour la dernière ligne, ils combinent les variables rétiniennes forme et taille.

Une première méthode consiste à combiner simplement les variables rétiniennes, lorsque les signes utilisent des variables rétiniennes différentes. Par exemple si l’on combine la couleur *rouge* et la forme *carré*, nous obtenons un *carré rouge*.

D’autres méthodes consistent à positionner les deux signes l’un par rapport à l’autre dans l’espace (voir exemple dans le tableau 5.2). Il est possible de les juxtaposer, avec ou sans recouvrement. La juxtaposition présente cependant un risque d’interférence avec les autres signes environnants, dans notre exemple s’il y a un autre signe placé près de l’une des extrémités de la flèche, le risque est grand d’associer la flèche avec ce signe plutôt qu’avec le rein. Il est aussi possible d’inclure l’un des signes à l’intérieur de l’autre. Dans ce cas, les deux signes ne jouent pas le même rôle : l’un est à l’intérieur et l’autre à l’extérieur. Le signe placé à l’intérieur masque une partie du signe extérieur, et le rend plus difficile à percevoir. Le signe extérieur doit donc être de préférence le plus simple des deux, selon les lois de ségrégation.

Règle 6 :

Lorsque l’on inclut un signe à l’intérieur d’un autre, le signe le plus simple doit être placé à l’extérieur.

Dans l’exemple du tableau 5.2, la flèche est une forme plus simple que le rein. Il est plus facile de reconnaître le rein dans la flèche, que la flèche dans le rein, et nous placerons donc la flèche à l’extérieur.

Certaines formes ne se prêtent pas à l’inclusion de signe à l’intérieur d’elles-mêmes, en particulier celles ayant des “trous” (voir par exemple notre pictogramme pour les poumons dans l’annexe G) ou étant trop complexes pour rester lisibles lorsqu’un signe est placé à l’intérieur (par exemple le rein). Afin de prendre en compte cela, nous avons choisi de décomposer la variable rétinienne “forme” en deux : “forme extérieure” et “pictogramme”. Les formes extérieures sont les formes qui peuvent recevoir un signe à l’intérieur d’elles-mêmes, tandis que les pictogrammes sont l’ensemble des autres formes. Il est donc possible de placer un pictogramme à l’intérieur d’une forme extérieure, mais pas l’inverse.

Le tableau 5.2 récapitule les différentes méthodes de combinaison de signe, et indique pour chacune la *séparabilité*, c’est à dire la propension à faciliter ou non la distinction entre les deux signes au sein de la combinaison produite.

Lorsque le nombre de signes à combiner est grand, notamment suite à l’utilisation d’une grammaire combinatoire, **il est important de hiérarchiser les signes afin de rendre explicite la structure**

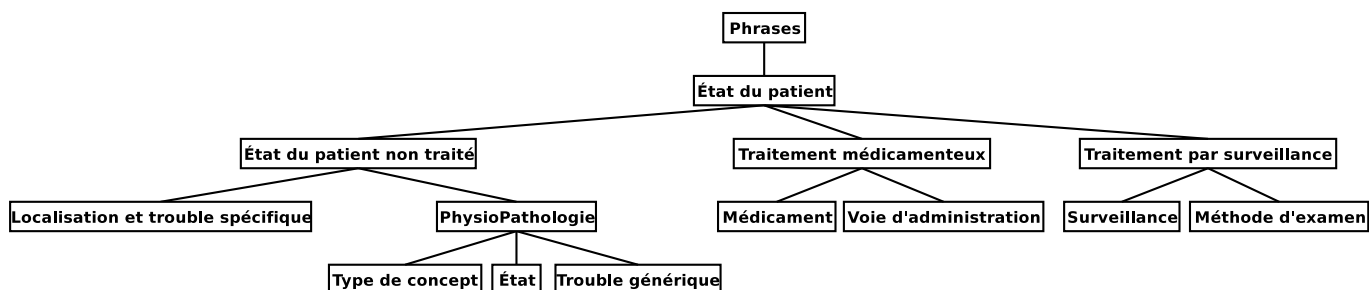


Fig. 5.4 – Décomposition des différents concepts et attributs en arbre (issu du modèle UML de la figure 4.4).

Les attributs localisation et trouble spécifique ont été regroupés ensemble car nous avons vu qu’il n’était pas possible de les représenter séparément.



Fig. 5.5 – Exemples de pictogramme à forme creuse (à gauche) et à forme pleine (à droite).

du modèle de connaissance sous-jacent. Pour cela, les signes ne doivent pas être combinés en une seule fois mais en plusieurs étapes : plusieurs petits groupes de signes représentant des attributs sémantiquement proches seront combinés ensemble, puis les signes résultant de ces combinaisons seront eux-mêmes combinés entre eux, et ainsi de suite.

Les schémas UML de la figure 4.4, qui ont été remis sous forme d’arbre en figure 5.4, permettent de déterminer quels attributs sont proches sémantiquement : les combinaisons de signes devront correspondre aux noeuds intermédiaires de l’arbre. Par exemple les attributs voie d’administration et médicament sont sémantiquement proches et seront combinés ensemble.

Règle 7 :

Les différents signes doivent être combinés en plusieurs fois, en suivant la structure arborescente du modèle de connaissance.

Lorsque le médecin regarde une phrase écrite en VCM, nous souhaitons qu’il décompose visuellement les différents signes contenus dans la phrase de la même manière que sur l’arbre de la figure 5.4. Pour cela, il faut que le médecin commence par décomposer selon les noeuds placés au sommet de l’arbre : la première décomposition doit être celle de la phrase en états du patient, suivie de la décomposition des états du patient en états du patient non traités et en traitements, et ainsi de suite. Cela peut être obtenu en jouant sur la propriété de séparabilité des différentes méthodes de combinaison de signe :

Règle 8 :

Plus les combinaisons de signes ont lieu à des niveaux élevés dans l’arbre représentant la structure des connaissances, plus il faut utiliser des méthodes de combinaison de signe facilitant la distinction entre les signes combinés (séparabilité élevée).

Enfin, nous pouvons remarquer une sorte de “parallélisme” dans l’arbre : les représentations des traitements médicamenteux et des traitements par surveillance suivent des structures similaires. Ce parallélisme doit se retrouver au sein du langage graphique, en représentant les attributs par les mêmes variables rétinienne, et en combinant ces variables rétinienne de la même manière.

Règle 9 :

Lorsque l’arbre représentant la structure des connaissances présente des branches parallèles, celles-ci devront être représentées graphiquement de la même manière.

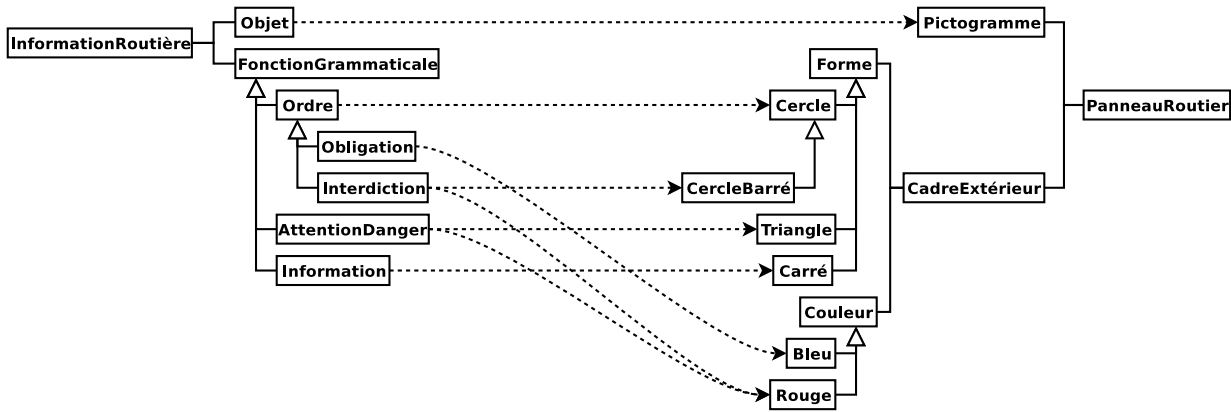


Fig. 5.6 – Exemple d’application de notre formalisme à la grammaire des panneaux routiers (version simplifiée).

5.2.5 Élaboration du lexique des pictogrammes

VCM comprend un certain nombre de pictogrammes. Afin de garder un ensemble homogène, le style graphique devra bien entendu être le même pour tous les pictogrammes. Un élément important à prendre en compte est le nombre de pictogrammes qui seront présents simultanément. En effet, si seule une dizaine de pictogrammes sont affichés, il est possible de les détailler et de les embellir avec des fioritures (par exemple des effets d’ombrage). En revanche, si une centaine de pictogrammes sont présents, ces fioritures sont sources de bruit visuel et compliquent inutilement une image d’ensemble qui est déjà compliquée.

Règle 10 :

Plus le nombre de pictogrammes présents simultanément est grand, plus les pictogrammes doivent être simples, et dépouillés de détails ou de “fioritures”.

En particulier, les pictogrammes à forme pleine sont plus simples que ceux à forme creuse (figure 5.5), car une forme creuse comprend deux contours, un pour délimiter l’extérieur du trait et un pour l’intérieur, alors que les formes pleines n’en ont qu’un seul. Notons que les panneaux routiers, ainsi que beaucoup de pictogrammes de signalisation, utilisent des formes pleines.

Sur le plan matériel, j’ai dessiné les pictogrammes moi-même, en utilisant le logiciel de dessin *The Gimp* (dessin dit “bitmap”). Les pictogrammes ont été réalisés en très haute résolution, au format PNG sur fond transparent, puis la résolution a été réduite pour l’affichage à l’écran.

5.2.6 Formalisation de la grammaire du langage graphique

De nombreuses méthodes ont été proposées pour formaliser la grammaire d’un langage, comme les formes de Backus-Naur ou les diagrammes de Conway. Cependant ces formalismes ne sont pas adaptés à la description de grammaires graphiques : elles définissent un “mot” comme une suite ordonnée de symboles, or dans un langage graphique les symboles ne suivent pas forcément un ordre linéaire ; nous avons vu dans le tableau 5.2 qu’il existe plusieurs manières de combiner des symboles graphiques et il n’y a pas “d’ordre” évident entre ces méthodes.

Afin de formaliser la grammaire d’un langage graphique, **nous proposons un formalisme utilisant le langage UML**, de la manière suivante : deux arbres de classes sont décrits l’un en face de l’autre ; celui de gauche représente les connaissances exprimées par le langage graphique, et celui de droite décrit les objets graphiques utilisés pour les représenter (pour VCM, l’arbre de gauche correspond à la figure 4.4). Ensuite, des relations de dépendance (flèches en pointillés en UML) sont utilisées pour indiquer quel élément de connaissance est représenté par quel élément graphique ; ces flèches peuvent se lire “est représenté par”. La figure 5.6 donne un exemple d’application de ce formalisme aux panneaux routiers.

5.3 Résultats

Le langage VCM a été construit de manière itérative, en appliquant la méthode décrite à la section précédente. Nous présenterons d’abord les résultats de l’analyse des combinaisons de localisations et de troubles génériques, et la grammaire du langage, puis nous donnerons un aperçu rapide du langage.

5.3.1 Analyse des combinaisons de localisations et de troubles génériques

Si l’on utilise une grammaire combinatoire pour représenter les pathologies comme des combinaisons (localisation, trouble générique), cela donne naissance à 700 combinaisons possibles, qui ont été classées manuellement en trois catégories (voir section 5.2.2) :

- 275 combinaisons utiles, soit 39%,
- 249 combinaisons cohérentes mais inutiles, soit 36%,
- 176 combinaisons incohérentes, soit 25%.

La grammaire combinatoire permet donc de représenter 275 pathologies avec seulement 55 signes de base (35 localisations + 20 troubles génériques), soit une diminution d’un facteur 5 du nombre de signes. Nous sommes donc dans une situation très favorable à la combinatoire.

5.3.2 Représentation des attributs

Le tableau 5.3 indique pour chaque attribut le nombre de valeurs qu’il peut prendre, les différentes variables rétinienne susceptibles de le représenter, et la variable qui a été retenue. La présence d’analogies ou de conventions existantes est également mentionnée.

Pour la représentation du type de concept et de l’état, nous n’avons pas voulu utiliser la variable pictogramme afin de limiter le nombre de pictogrammes utilisés ; sinon, il nous aurait fallu 3 pictogrammes pour représenter seulement “pathologie cardiaque en cours” (un pour “en cours”, un pour “pathologie” et un pour “cardiaque”) ! De plus, l’utilisation de la variable pictogramme, qui peut représenter un très grand nombre de valeurs, ne se justifie pas pour ces deux attributs. La couleur a donc été choisie pour représenter le type de concept, en utilisant des valeurs conventionnelles pour certains types de concept, et des valeurs arbitraires pour les autres.

Le risque est souvent représenté par un triangle (par exemple dans les panneaux routiers). Cependant, ce triangle consomme beaucoup de place, et surtout il ne satisfait pas à la règle de représentation des relations *est-un* (voir section 5.2.3).

De même, des formes extérieures arbitraires ont été choisies pour représenter l’état pathologique ou physiologique. Nous avons choisi le carré pour les états pathologiques, car cette forme se prête bien à des “extensions” ou “modifications” correspondant à des troubles génériques qui viendraient préciser la pathologie (par exemple le carré peut être transformé en flèche). Au contraire, les états physiologiques n’ont pas besoin d’être précisés. Sur l’ensemble des attributs, nous avons seulement 5 signes arbitraires en tout (2 formes : rond et carré, et 3 couleurs : rouge, marron et bleu).

La représentation analogique des localisations anatomiques par des couleurs n’est pas possible car la longueur de la variable rétinienne couleur ne permet pas de représenter un nombre suffisant de localisations. Les troubles spécifiques ont été combinés avec les localisations car nous avons vu que, pour ces deux attributs, il n’était pas intéressant d’utiliser la combinatoire.

L’attribut trouble générique a été divisé en quatre, car la variable rétinienne la mieux adaptée n’est pas la même pour les différents types de troubles génériques. Pour les variations, la représentation analogique consistant à faire varier le nombre de pictogrammes est séduisante ; elle permettrait de représenter par exemple les polyglobulinémies par un vaisseau sanguin avec un nombre élevé de globules rouges. En revanche, il est beaucoup plus difficile de représenter l’anémie avec cette méthode. C’est pourquoi nous avons préféré utiliser les flèches.

L’attribut sexe peut être représenté soit par des couleurs, soit par des pictogrammes en utilisant des conventions connues ; nous avons choisi les pictogrammes car les autres caractéristiques du patient

attribut	nombre de valeurs	variables rétinienne possibles, longueur				
		forme, 10	pictogramme, infini	couleur, 10		
type de concept	4		pictogramme suggérant le concept : état actuel : personne, antécédent : calendrier, médicament : caducée,... (A), risque : triangle (C)	état actuel : rouge (O), risque : orange (C), antécédent : marron (O), médicament : vert (C), surveillance : bleu (O)	orientation, 12	
localisation et trouble spécifique	35+50	(-)		pictogrammes représentant l'organe (A) + le trouble associé s'il est présent : Alzheimer : SNC + trou (A), vomissement : estomac + flèche sortant (C),...	couleurs des organes : coeur/sang : rouge, SNC : gris,... (A-)	(-)
état (pathologique ou non)	2	physiologique : rond, pathologique : carré (O)		physiologique : personne debout, pathologique : personne alitée (A)		
trouble	3		pictogramme où le nombre d'éléments varie : hyperglobulinémie : 3 globules rouges, normal : 2 globules, anémie : 1 globule (A)			flèches vers le haut ou le bas (C)
générique	6		pictogrammes des agents infectieux (A)			
proc. macroscopique	4	forme suggérant le trouble : douleur : explosion (C), inflammation : flamme (A),...				
proc. microscopique	4		pictogrammes indiquant le processus : atteinte nerveuse : cellule nerveuse,... (A)		atteinte vasculaire : rouge, atteinte nerveuse : gris,... (A)	
voie d'administration	3		pictogrammes d'une forme médicamenteuse associée à la voie d'administration (A)			
méthode d'examen	3		pictogrammes d'un objet suggérant le type d'examen : biologie : tube à essai,... (A)			
âge	4		pictogrammes suggérant les classes d'âge : nourrisson : bébé + berceau,... (A)			
sexe	2		homme : ♂, femme : ♀ (C)		homme : bleu, femme : rose (C)	
mode de vie	6		pictogrammes des modes de vie : tabac : cigarette,... (A)			
action du médecin	5		pictogrammes des actions : changer la posologie : gobelet doseur,... (A)			

Tab. 5.3 – Nombre de valeurs et variables rétinienne possibles pour chaque attribut.

La variable retenue est soulignée; (A) indique une analogie, (C) une convention existante, (O) un choix arbitraire et (-) l'impossibilité d'utiliser cette variable, car sa longueur est trop faible.

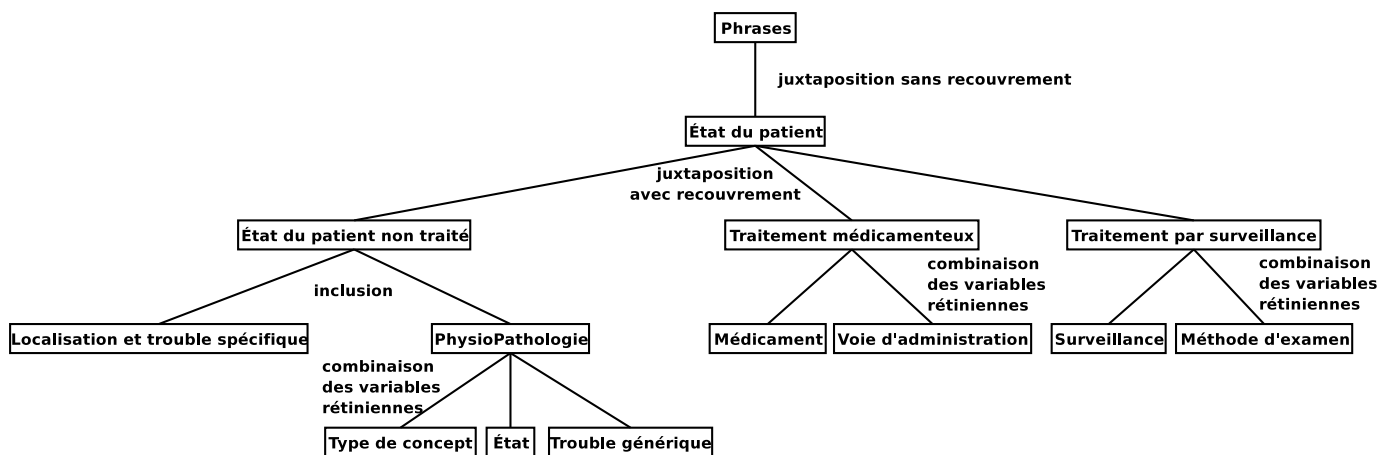


Fig. 5.7 – Décomposition des différents concepts et attributs en arbre (issu du modèle UML de la figure 4.4).

Pour chaque noeud intermédiaire, la méthode de combinaison de signe utilisée est indiquée.

(âge, mode de vie) étaient déjà représentées de cette manière.

5.3.3 Représentation des concepts

Les concepts médicaux sont représentés en combinant les signes représentant les attributs du concept, selon la grammaire combinatoire. La figure 5.7 montre les méthodes de combinaison de signe que nous avons utilisées, pour les différents niveaux de l'arborescence du modèle de connaissance. Les icônes représentant les états du patient ont été décomposées en trois parties : une première partie représentant la localisation et éventuellement le trouble spécifique associé, une seconde avec les autres éléments de physiopathologie (état pathologique ou non, trouble générique, type de concept) et une troisième avec le traitement correspondant (médicament ou surveillance). La première partie est représentée par un pictogramme et la seconde partie par une forme extérieure ; le pictogramme est inclus à l'intérieur de cette forme extérieure.

Les pictogrammes indiquant l'étiologie ont été placés à gauche car l'étiologie indique la cause, laquelle est normalement placée à gauche, c'est à dire avant, la conséquence. Les pictogrammes indiquant le mécanisme de la pathologie ont été placés à droite, pour faciliter la distinction avec les pictogrammes d'étiologie.

Enfin, le traitement associé est représenté par un autre signe, qui est combiné au signe précédent par juxtaposition avec recouvrement. Nous avons choisi de le placer "en exposant", c'est-à-dire en haut à droite. Ce choix est en partie arbitraire, le seul élément pris en compte étant de ne pas recouvrir d'élément important, comme les pictogrammes d'étiologie ou de mécanisme pathologique.

La figure 5.8 décrit la grammaire combinatoire en langage UML (voir section 5.4 pour une description du formalisme utilisé). La grammaire du langage VCM a aussi été formalisée en utilisant des BNF (*Backus-Naur Forms*), cependant cette formalisation est beaucoup moins lisible que celle que nous proposons ici. La BNF du langage VCM est joint en annexe H.

5.3.4 Représentation des phrases

Pour la représentation des phrases VCM, nous avons besoin de représenter des notions abstraites comme la négation l'impératif,... qui ne peuvent pas toujours être représentées par analogie. En revanche, il existe un certain nombre de conventions, bien connues ou issues en particulier de la signalisation routière.

Au sein des phrases VCM, nous avons distingué 6 notions ; certaines sont partagées par plusieurs éléments des phrases, par exemple la notion d'ordre est utilisée pour les actions à faire et les actions à ne pas faire. Le tableau 5.4 liste les notions ainsi que les analogies et conventions existantes correspondantes, et celles que nous avons retenues. Pour représenter les relations si-alors, nous avons

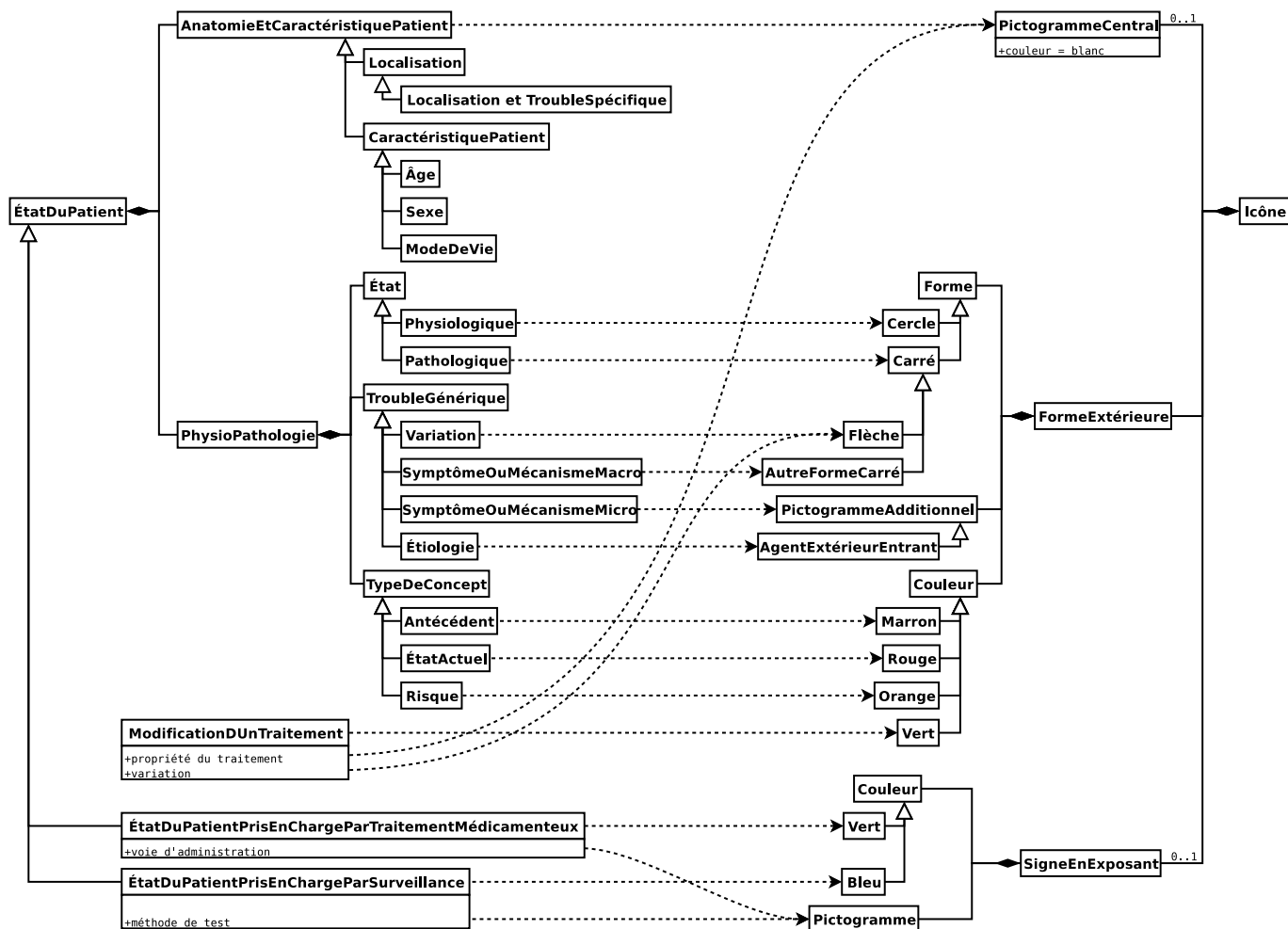


Fig. 5.8 – Représentation formelle en UML de la grammaire et de la sémantique des icônes du langage VCM.

Les flèches de dépendance (en pointillés) signifient “est représenté par”.

notion	éléments concernés	analogie	conventions
relation si-alors	condition	-	<u>flèche</u> : condition -> conséquence point d'interrogation : condition ? conséquence
affirmation	affirmation	<u>signe représentant la chose affirmée</u> , sans modification	
impératif (ordre)	action à faire, action à ne pas faire	-	<u>cercle</u> (dans la signalisation routière)
négation	action à ne pas faire	-	<u>barré</u>
ET logique	condition, affirmation, action à faire, action à ne pas faire	<u>juxtaposition des opérandes</u>	symbole & séparant les opérandes
OU logique	condition, action à faire	<u>dérivation</u> : un “chemin” se sépare en plusieurs voies parallèles; les opérandes sont chacune placées sur un “chemin” différent (cf les circuits électroniques, par exemple)	<u>barre verticale</u> séparant les opérandes

Tab. 5.4 – Analogies et conventions existantes pour représenter les notions présentes dans les phrases VCM. Un “-” indique qu’il n’est pas possible d’utiliser l’analogie pour cette notion. Les passages soulignés correspondent aux modes de représentation que nous avons retenus.

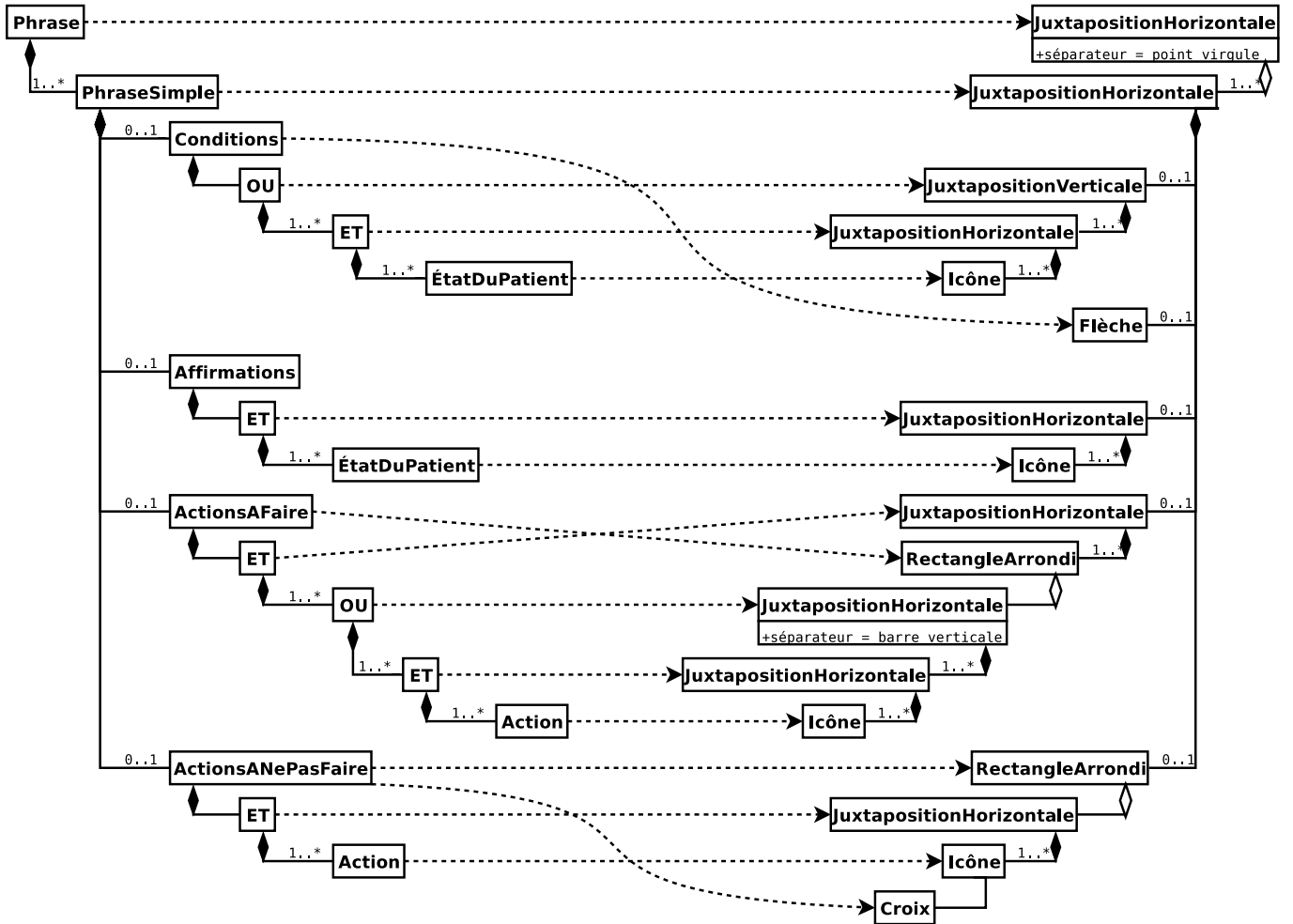


Fig. 5.9 – Représentation formelle en UML de la grammaire et de la sémantique des phrases du langage VCM.

Les flèches de dépendance (en pointillés) signifient “est représenté par”.

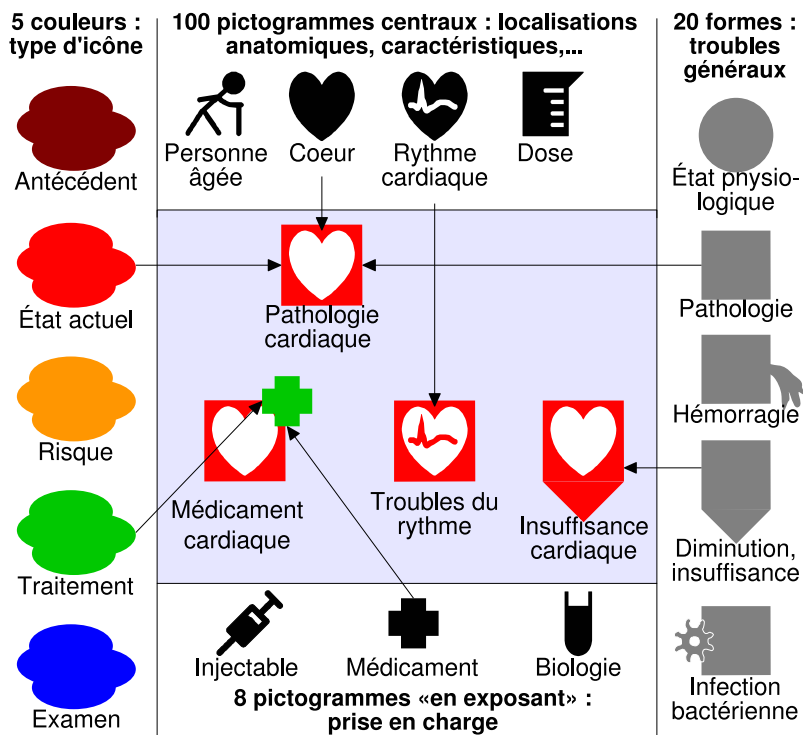


Fig. 5.10 – Construction des icônes du langage VCM.

préféra la flèche au point d'interrogation, car ce dernier est un caractère "textuel". Nous avons choisi d'entourer les ordres par un rectangle arrondi plutôt que par un cercle, afin de consommer moins de place. Le symbole "&" est un symbole arbitraire et nous avons préféré l'éviter. Enfin, pour les relations OU logiques, nous avons retenus les dérivations pour les OU dans les conditions, et la barre verticale pour les relations OU dans les actions à faire. Ce choix est justifié par le fait que la flèche (qui suit les conditions) crée un "chemin" propice à la dérivation, qui n'existe pas dans les actions à faire.

La figure 5.9 décrit la grammaire des phrases en langage UML (voir section 5.4 pour une description du formalisme utilisé). Cette grammaire a aussi été formalisée en BNF, en annexe H.

5.3.5 Dessin des pictogrammes

Lors de l'analyse de TAL du corpus de RCP (voir section 4.2.3.2 et 4.3.2), nous avons trouvé un total de 19522 concepts médicaux (= 18322 états pathologiques + 1200 états non pathologiques) sur les 278 RCP, soit une moyenne de 70 concepts par RCP. Puisque VCM représente chaque concept par une icône, cela correspond à une moyenne de 70 icônes par RCP, ce qui est un nombre assez important. Les pictogrammes VCM devront donc être le plus simple possible.

J'ai réalisé des pictogrammes pour les différentes localisations anatomico-fonctionnelles, les étiologies, les caractéristiques du patient et habitudes de vies. La version actuelle de VCM comprend 120 pictogrammes et formes différents (voir annexe G).

5.3.6 Présentation du langage VCM version 1.0

Nous donnons ici une brève présentation du langage, dans sa version finale 1.0. Pour la présentation complète, il faut se référer à l'annexe E; cette annexe est aussi disponible en ligne à l'adresse suivante : http://vcm.limbio-paris13.org/guide_vcm/vcm_fr.pdf et en anglais : http://vcm.limbio-paris13.org/guide_vcm/vcm_en.pdf. Plusieurs versions préliminaires ont été présentées lors de congrès ou de publications [85, 86, 87, 88].

5.3.6.1 Représentation des icônes

VCM comprend 5 couleurs et 120 pictogrammes (figure 5.10) qui sont utilisés pour construire les icônes du langage. Les icônes représentent les différents concepts médicaux ; chaque type de concept étant associé à une couleur :

- l'état actuel du patient (incluant pathologies, symptômes, états physiologiques, mode de vie et caractéristiques comme l'âge) en **rouge**,
- les risques de pathologies et de symptômes en **orange**,
- les antécédents de pathologie en **marron**,
- les traitements en **vert**,
- les surveillances en **bleu**.

Les états non pathologiques actuels du patient (états physiologiques, mode de vie et caractéristiques) sont représentés dans un rond **rouge**, et l'état est précisé par un pictogramme blanc à l'intérieur, par exemple :



Personne âgée

Les états pathologiques actuels du patient, les risques et les antécédents sont représentés par un carré de la couleur correspondante, et la localisation anatomico-fonctionnelle est représentée par un pictogramme blanc à l'intérieur, par exemple :



Pathologie cardiaque



Risque d'atteinte rénale



Antécédents de pathologie pulmonaire

Pour les pathologies, il est possible de préciser le type de trouble de deux manières différentes :

- Soit en modifiant le pictogramme, lorsque **le trouble est spécifique de la localisation anatomico-fonctionnelle**, par exemple :



Troubles du rythme (spécifique du coeur)

- Soit en complétant le carré par des modificateurs de forme (prolongement du carré par une flèche indiquant une variation du niveau d'activité d'une fonction, changement de la forme extérieure du carré indiquant un symptôme ou mécanisme général, ou bien ajout d'un petit pictogramme additionnel indiquant une étiologie ou une localisation systémique), lorsque **le trouble est générique** et peut s'appliquer à plusieurs localisations anatomico-fonctionnelles, par exemple :



(

Insuffisance cardiaque (prolongement du carré par d'une flèche vers le bas : hypofon-



ction)
Insuffisance rénale



Angor (changement de la forme extérieur en explosion : douleur)



(

Infection rénale bactérienne (ajout d'un pictogramme additionnel : étiologie bacté-



rienne)
(

Pathologie ischémique (ajout d'un pictogramme additionnel : localisé au système

circulatoire)

Les médicaments sont représentés par l'icône de la pathologie qu'ils traitent, à laquelle on ajoute une petite croix verte en haut à droite, par exemple :



Anti-arythmiques



Cardiotoniques (traitements de l'insuffisance cardiaque)

Cette croix verte peut être remplacée par un pictogramme représentant un tube de pommade, un comprimé ou une seringue pour préciser la voie d'administration (topique, orale ou parentérale), notamment lorsque cela permet de distinguer des classes thérapeutiques :



Anti-asthmatiques locaux



Anti-asthmatiques systémiques



Anti-diabétiques oraux



Insulinothérapie

Les surveillances sont représentées par l'icône du risque surveillé, auquel on ajoute un petit pictogramme bleu en haut à droite, qui précise le type d'examen (un stéthoscope pour examen clinique, un signal pour examen fonctionnel, un tube à essai pour la biologie ou un rayonnement pour l'imagerie), par exemple :



Surveillance clinique du coeur



Surveillance fonctionnelle du rythme cardiaque (ECG)



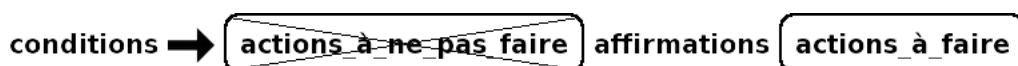
Surveillance rénale biologique (par exemple clairance)



Surveillance des os par imagerie (par exemple radiographie)

5.3.6.2 Représentation des phrases

Le langage VCM permet ensuite de combiner plusieurs icônes ensemble pour former des phrases, en suivant une grammaire particulière. La grammaire des phrases correspond au modèle suivant :



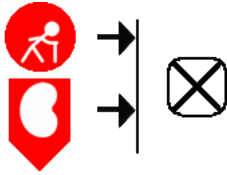
Dans ce modèle, “conditions”, “affirmations”, “action à faire” et “actions à ne pas faire” sont remplacés par les icônes des termes correspondants. Lorsque l'on représente un RCP, la prise du médicament du RCP est une condition implicite qui n'est pas représentée afin d'alléger les phrases ; ainsi l'on ne dira pas “ce médicament est contre-indiqué avec l'asthme” mais “contre-indiqué avec l'asthme”. Enfin, un carré arrondi barré vide (symbole “action à ne pas faire” sans icône à l'intérieur) signifie “ne pas prescrire”.

5.3.6.3 Exemples

Voici deux exemples (qui reprennent ceux de la figure 4.5) :



= En cas de troubles du rythme, il est conseillé d'effectuer un ECG.



= (Ce médicament est) contre-indiqué chez la personne âgée *et* en cas d'insuffisance rénale (le “*et*” a en fait la valeur d'un OU logique).

5.4 Discussion et conclusion

5.4.1 À propos de la méthode

Nous avons vu que la méthode de conception des pictogrammes proposée par l'ISO n'était pas satisfaisante (voir section 2.2.2.1 et 2.2.5). En effet elle ne permet pas l'utilisation d'une grammaire, or nous avons vu que la grammaire combinatoire était très intéressante dans un langage comme VCM. De plus, la méthode de l'ISO repose sur un groupe d'experts du domaine ; si ceux-ci connaissent leur domaine, ils sont en général assez ignorants des aspects cognitifs, et aussi artistiques, qui sont pourtant importants dans un langage graphique. Or nous avons vu (à la section 2.1.6) que la conception d'un langage graphique intuitif n'est pas une tâche intuitive.

La méthode que nous avons proposée dans ce chapitre présente l'avantage de prendre en compte les règles cognitives, et de permettre la conception de grammaire combinatoire. L'aspect “combinatoire” ou “compositionnel” des concepts médicaux a déjà été souligné dans la littérature [126].

En ce qui concerne les relations entre les concepts, nous avons choisi de nous limiter à la représentation des relations *est-un*. En effet, celles-ci jouent un rôle important en médecine. Elles peuvent être représentées graphiquement de manière simple, de sorte à être perçues sans raisonnement : si l'oeil voit le signe qui représente A, alors c'est un A, sinon ce n'est pas un A. Des relations plus complexes (appartient à, traite,...) peuvent être représentées graphiquement, par exemple à l'aide de flèches comme nous l'avons fait pour les relations si-alors, cependant mon opinion est qu'elles ne pourraient pas être perçues aussi rapidement que les relations *est-un* graphiques, car leur interprétation nécessite une phase de raisonnement qui ne peut pas se faire au premier niveau d'interprétation (voir figure 2.6). Ce point pourrait faire l'objet d'une étude dans le domaine des sciences cognitives, mais sort du cadre de cette thèse.

Afin de formaliser la grammaire graphique de VCM, nous avons proposé un formalisme utilisant le langage UML (voir section 5.2.6). Comparé aux formalismes existants, il présente l'avantage d'être graphique tout en n'ayant pas les limitations habituelles des formalismes graphiques (voir section 2.2.5), grâce à l'utilisation du langage UML.

5.4.2 À propos des résultats

Le langage VCM représente de manière explicite une partie importante des informations qui étaient implicites dans le langage textuel. Nous avons vu que cela facilitait l'apprentissage du langage, mais accélérerait aussi les recherches visuelles avec ce langage. Chaque icône VCM est alors une définition du concept qu'elle représente, à la manière d'une ontologie. Les langages graphiques pourraient donc être une approche intéressante pour la représentation d'ontologies. En effet, un certain nombre d'ontologies médicales sont disponibles ; cependant celles-ci sont souvent trop complexes pour

être utilisées en clinique. Cependant elles pourraient être utiles, tant pour les professionnels de santé que pour les patients, pour lesquels l'accès aux définitions des concepts médicaux est important car ils n'ont pas de formation médicale et ne connaissent pas ces concepts.

5.4.3 Conclusion

À partir des aspects cognitifs que nous avons étudiés au chapitre 2, nous avons mis au point des éléments de méthode pour concevoir un langage graphique facile à apprendre et rapide à lire, en rendant explicites un maximum d'informations implicites. Ensuite nous avons conçu le langage VCM en nous appuyant sur le modèle de connaissance que nous avons construit au chapitre 4. Ce langage représente de manière explicite des informations qui sont implicites dans les textes médicaux, et nous avons fait l'hypothèse que cela permettrait de faciliter l'apprentissage du langage et d'accélérer sa lecture. Nous allons maintenant voir comment appliquer des méthodes de visualisation d'information à ce langage.

Chapitre 6

Utilisation de VCM dans les textes médicaux et application de techniques de visualisation d'information

6.1 Introduction et objectifs

En introduction, nous avons vu que l'objectif du langage VCM n'était pas de remplacer les textes médicaux car VCM ne pouvait pas être suffisamment précis. En conséquence, nous devons mettre au point des méthodes pour combiner les textes médicaux avec le langage VCM. Ces méthodes pourront utiliser des techniques de visualisation d'information et de *Fisheye*, afin de rendre la visualisation interactive. Le langage VCM pourra servir de "filtre" afin de permettre à un professionnel de santé de faire le tri entre les connaissances qui l'intéressent, c'est-à-dire celles qui s'appliquent à son patient, et les autres. Si le médecin est intéressé, il pourra lire le texte correspondant pour avoir plus de détails, par exemple en cliquant sur l'icône pour l'afficher (voir figure 1.1).

La mise au point des méthodes de visualisation s'est faite de manière itérative, en s'appuyant sur le développement d'un prototype de base de connaissance sur le médicament mettant en oeuvre ces méthodes. Ce prototype sera présenté dans la section résultat.

6.2 Matériel et méthodes

6.2.1 Matériel pour la réalisation du prototype

6.2.1.1 Langage de programmation

Le prototype a été développé en Python, un langage de programmation orienté objet, multi-plateforme et disponible en logiciel libre (voir <http://python.org>). Le prototype se compose d'un ensemble de scripts générant des pages web (HTML) dynamiques (Javascript) et des images (PNG) à partir des données; il peut aussi fonctionner à la manière d'un serveur web. Les pages HTML sont ensuite visualisées dans un navigateur Internet standard (les navigateurs Mozilla Firefox et Internet Explorer ont été testés).

6.2.1.2 Données

Le prototype utilise des données fournies sous forme d'un fichier XML (*eXtensible Markup Language*; un fichier par document à mettre en VCM). J'ai développé ce format XML spécialement pour faciliter la traduction de document textuel en VCM; la description du format (DTD) est donnée en annexe J. Le fichier XML correspond à une monographie ou à un RPC dans lequel les mots ou passages ayant un sens médical ont été balisés. Par exemple, "ce médicament est contre-indiqué en cas d'hypertension" devient :

```
<phrase mod="npsm :phrase_contre_indication">
ce médicament est contre-indiqué en cas d'
<terme code="cim10 :I10" mod="npsm :condition">hypertension</terme>
</phrase>
```

Ce format XML utilise les classifications CIM10 et ATC pour coder les pathologies et les médicaments, ainsi que la classification NPSM (Non-Pathologie Non-Médicament) que j'ai développée afin de coder tout le reste (fonctions grammaticales, type de concept, caractéristiques patient,... voir la description de cette classification en annexe I).

Il existe deux moyens pour obtenir de tels fichiers XML :

- **Partir d'un RCP en "texte brut"** (que l'on peut trouver sur le site web de l'AFSSAPS par exemple) **et le baliser à la main**. Cette opération est relativement longue et complexe, car il faut baliser et coder les différentes parties du RCP (par exemple, baliser "hypertension" comme étant un nom de pathologie et lui attribuer le code CIM10 I10 : "hypertension essentielle (primitive)"). Cette étape fastidieuse pourrait être facilitée par des outils de balisage et de codage semi-automatiques ou automatiques ; cependant de tels outils sortent du cadre de cette thèse.
- **Partir de données déjà codées dans une base de connaissances médicamenteuses existantes**. J'ai réalisé un script Python qui permet de générer un fichier XML à partir des données de Thériaque pour une spécialité donnée.

6.2.2 Juxtaposition du texte et des icônes

Une première méthode pour combiner les icônes du langage VCM avec les textes médicaux consiste simplement à **placer les icônes VCM à côté des passages correspondants dans le texte**. Afin d'améliorer la visibilité de l'ensemble, il est conseillé de réserver un espace pour les icônes, en marge du texte, à gauche ou à droite. Le médecin peut alors passer rapidement en revue les icônes placées dans la marge, et lire le texte correspondant uniquement si celui-ci l'intéresse.

Le résultat est un texte enrichi par des icônes, et s'apparente donc à un texte. Dans l'état de l'art, nous avons vu plusieurs techniques de visualisation d'information qui appliquent aux textes le principe du *Fisheye* (section 2.3.2). La technique du repliement concerne les textes structurés. Cette technique s'applique bien aux RCP car ils suivent un plan bien précis. Le repliement permet ainsi de choisir rapidement quelles sections du RCP sont visibles ; cela est d'autant plus intéressant qu'un certain nombre de sections ne sont que rarement utilisées en clinique (informations administratives, pharmaco-cinétique,...).

6.2.3 Représentation à l'aide d'un schéma anatomique

La simple juxtaposition du texte et des icônes ne permet pas réellement de réduire le volume d'information : lorsque le texte est long, le nombre d'icônes peut devenir très important. Nous souhaitons à présent mettre au point une méthode de visualisation utilisant le *Fisheye* : c'est-à-dire qui n'affiche qu'une partie des connaissances (en l'occurrence les icônes VCM), tout en affichant la totalité des connaissances (le texte) sur une zone "focus" qui a été sélectionnée par le médecin, par exemple en cliquant sur les icônes (voir section 2.3.1). Cette représentation interactive sera donc entièrement graphique et suffisamment synthétique pour tenir dans un seul écran.

En ce qui concerne les contre-indications, interactions médicamenteuses et précautions d'emplois en garde d'un RCP, celles-ci ne s'appliquent que chez certains patients : par exemple la précaution d'emploi "En cas de diabète, diminuer la posologie" ne concerne que les patients diabétiques. Dans un premier temps, seules les conditions de la phrase (ici, "En cas de diabète") vont intéresser le médecin ; la suite de la phrase ("diminuer la posologie") pourra être affichée à la demande, lorsque le médecin cliquera sur l'icône du diabète.

Pour les effets indésirables, ce sont les risques associés qui sont importants ("Risque de vomissement",...). Lorsque le médecin cliquera sur l'icône du risque, les autres informations disponibles

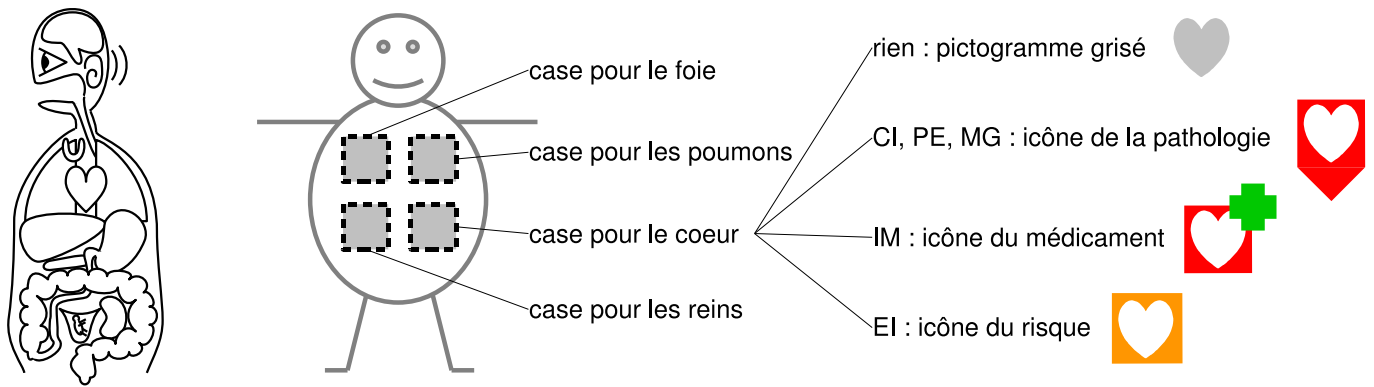


Fig. 6.1 – Exemples de schémas anatomiques.

Le schéma de gauche est réaliste, tandis que celui de droite est simplifié et découpé en “case” ; pour la case du coeur, les différents éléments possibles sont indiqués.

seront affichées, comme les actions à faire pour prévenir l’effet indésirable.

Que ce soit pour les conditions des contre-indications, interactions médicamenteuses et précautions d’emploi-mises en garde, ou les risques des effets indésirables, **les professionnels de santé attachent une grande importance à la localisation anatomico-fonctionnelle** [20] ; par exemple dans “insuffisance cardiaque” le mot “cardiaque” a plus d’importance que le mot “insuffisance”. Lorsque la localisation n’est pas précisée, c’est l’étiologie qui devient prépondérante dans le raisonnement médical, en particulier pour les infections. Il paraît donc logique de classer les contre-indications, interactions médicamenteuses, précautions d’emploi-mises en garde et effets indésirables en fonction de la localisation ou de l’étiologie de l’état physiopathologique concerné (pour les contre-indications, les précautions d’emploi et les mises en garde), de la pathologie traitée par le médicament (pour les interactions médicamenteuses) ou du risque (pour les effets indésirables).

Les différentes localisations et étiologies pourront donc être placées sur un schéma anatomique ; celui-ci peut soit être réaliste, soit simplifié et organisé en “cases” fixes et semi-arbitraires, correspondant chacune à une localisation ou à une étiologie (figure 6.1). Le schéma simplifié est préférable pour les raisons suivantes :

- Il permet de représenter clairement chaque organe, en lui allouant une case de taille fixe, tandis que sur le schéma réaliste, il est difficile de représenter simultanément les organes de certaines régions comme l’abdomen, à cause de la densité en organes et de leurs positions dans l’espace.
- Les cases ont toutes la même taille, alors que sur un schéma réaliste, les organes ont des tailles différentes, ce qui peut conduire à des interprétations erronées. Par exemple, une contre-indication sur les poumons, organes volumineux, risque d’être interprétée comme plus importante qu’une contre-indication sur la thyroïde, car celle-ci sera représentée plus petite.
- Il est possible de placer des icônes VCM dans les cases, ce qui n’est pas possible avec une représentation réaliste.

Cependant, certaines précautions d’emploi n’ont pas de condition (par exemple “Surveiller la fonction rénale”) ; dans ce cas il faudra trouver un autre moyen de les représenter.

Par ailleurs, nous ferons la même hypothèse que lors de la conception du langage VCM : il est possible d’accélérer l’accès aux connaissances en représentant de manière explicite des connaissances qui sont implicites dans les textes (Hypothèse 1, voir section 5.2). Dans un RCP, **l’absence de contre-indication, d’interactions médicamenteuses, de précautions d’emploi, de mises en garde ou d’effets indésirables concernant une localisation ou une étiologie donnée est une information implicite** : par exemple le texte ne mentionne jamais explicitement “Ce médicament *n’est pas* contre-indiqué en cas de pathologie cardiaque”. Pour en prendre connaissance le médecin est obligé de parcourir la totalité du texte, ce qui prend du temps.

Nous proposons de représenter explicitement sur notre schéma l’absence de d’information pour une localisation ou une étiologie, en faisant figurer dans la case correspondante un symbole particulier, par exemple le pictogramme de la localisation ou de l’étiologie en grisé (hors icône).

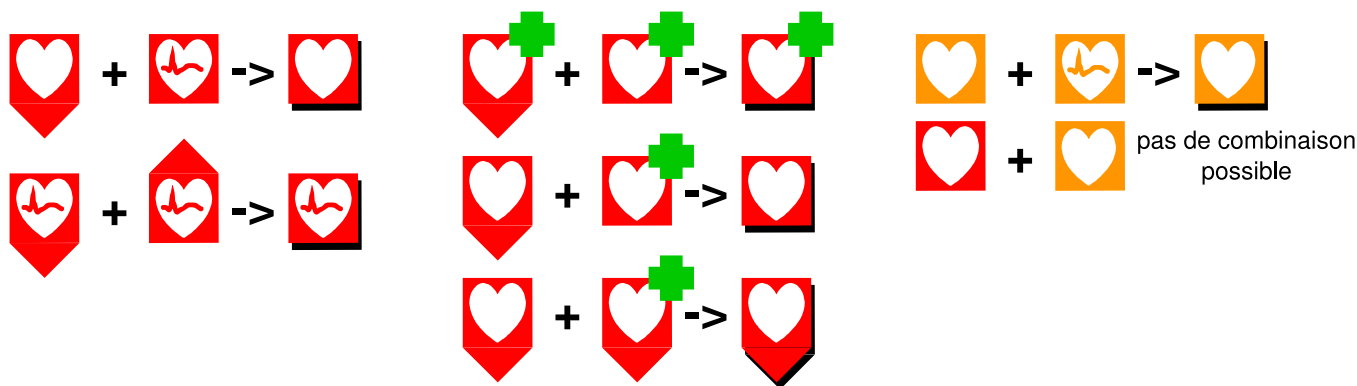


Fig. 6.2 – Exemples de combinaisons d’icônes.

6.2.4 Règles de combinaison d’icônes

Lors de la représentation sur un schéma anatomique, il est nécessaire de combiner plusieurs icônes ensemble, lorsque ces icônes devraient être placées dans la même case. Pour combiner les icônes entre elles, les relations *est-un* issues des classifications CIM10 ou ATC ne peuvent pas être utilisées car celles-ci présentent des lacunes, par exemple la CIM10 ne possède pas de terme “troubles du rythme”. En revanche la grammaire combinatoire permet de combiner facilement les icônes de deux concepts : l’icône résultante est une icône correspondant à un concept “virtuel” composé seulement des attributs communs aux deux concepts sources (voir exemples figure 6.2). Afin d’indiquer que l’icône ne représente pas un seul concept, mais la combinaison de plusieurs, une ombre noire est ajoutée derrière l’icône. Cette ombre donne l’image d’un “empilage” de plusieurs icônes.

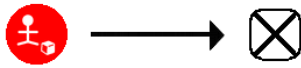
Les relations *est-un* que nous avons définies entre le patient traité pour l’état X et le patient dans l’état X d’une part, et le patient surveillé pour l’état X et le patient dans l’état X d’autre part, peuvent être utilisées lors de la combinaison des icônes : ainsi l’icône des médicaments cardiaques pourra être combinée avec celle des pathologies cardiaques. Il est donc possible de combiner les interactions médicamenteuses avec les contre-indications et les précautions d’emploi-mises en garde, et de les représenter ensemble sur un même schéma anatomique.

En revanche, les relations de notre modèle ne permettent pas de combiner un risque avec un état actuel. Il n’est donc pas possible de combiner les effets indésirables avec les contre-indications, interactions médicamenteuses et précautions d’emploi-mises en garde. Les effets indésirables devront donc être représentés sur un schéma différent.

6.2.5 Interaction médicamenteuse avec l’automédication

Certaines interactions médicamenteuses concernent des médicaments sujets à l’automédication, comme les AINS (Anti-Inflammatoires Non-Stéroïdiens, dont l’aspirine). Cette information est importante car, le patient étant susceptible de prendre ces médicaments de lui-même, le médecin ou le pharmacien doivent lui signaler l’interaction. Cependant, cette information est rarement donnée de manière explicite dans le RCP : c’est au médecin ou au pharmacien de lire l’ensemble des interactions médicamenteuses et d’en déduire celles qui peuvent poser problème avec l’automédication. Or cela prend du temps, et médecin comme pharmacien peuvent ne pas y penser. C’est pourquoi il peut être intéressant d’indiquer de manière explicite les interactions médicamenteuses avec l’automédication, dans un outil d’aide à l’éducation du patient (voir section 3.7).

Puisque l’automédication se concentre principalement sur certaines classes de médicament, comme les antalgiques, j’ai considéré que les médicaments susceptibles de faire l’objet d’une automédication étaient les médicaments en vente libre (=hors liste) et appartenant à l’une des classes ATC suivantes : A01, A02, A03, A04, A06, A07, A11, A12, A16, D1, D2, D3, D4, D7, D8, D9, D10, D11, G02B, G03A, M01, M02, M03, N02, N05B, N05C, N06A, P03, R01, R02, R05, R06, R07, V06. Ces classes contiennent des médicaments des voies digestives (anti-diarrhéique,...), des traitements dermatologiques (désinfectant,...), les contraceptifs à usage topique ou par voie orale, des anti-inflammatoires et



4.2 Posologie et mode d'administration

Voie orale.
 Réservé à l'adulte.
 2 à 3 gélules par jour à répartir dans la journée, à prendre avec un grand verre d'eau.
 La prise doit être effectuée de préférence aux mêmes heures.

4.3 Contre-indications

(...)

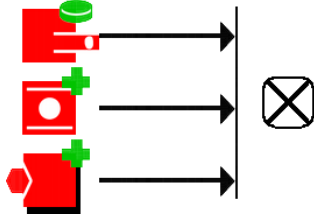
4.4 Mises en garde spéciales et précautions particulières d'emploi

(...)

4.5 Interactions avec d'autres médicaments et autres formes d'interactions

I) Interactions pharmacocinétiques

A) Associations contre-indiquées



* Médicaments anti-vitamine K
 * ciclosporine,
 * Antirétroviraux inhibiteurs de protéases et inhibiteurs non nucléosidiques de la transcriptase inverse.

Fig. 6.3 – Copie d'écran de l'outil de lecture, pour l'Arkogélule Millepertuis (un antidépresseur).

des antalgiques (dont l'aspirine), des anxiolytiques, des antidépresseurs, des anti-parasitaires (répulsif anti-poux,...), des traitements de la toux, des suppléments nutritionnels,...

6.3 Résultats

J'ai réalisé un prototype mettant en oeuvre le langage VCM et les différentes techniques de visualisation d'information proposées à la section précédente. Ce prototype comprend un outil de lecture du texte des RCP, un outil de vérification des contre-indications et des précautions d'emploi, un outil d'aide à l'éducation du patient et un outil d'aide au suivi thérapeutique (voir la classification des outils section 3.7). Ces trois derniers outils s'appuient sur "Monsieur VCM", un schéma anatomique interactif.

6.3.1 Outil de lecture

L'outil de lecture (figure 6.3) permet de lire l'ensemble du texte du RCP ; il utilise la juxtaposition pour combiner le texte et les icônes VCM. Celles-ci sont placées dans une marge à gauche du texte. Ce mode de visualisation reste proche des bases de connaissances médicamenteuses traditionnelles (type Vidal électronique). Cependant il s'appuie sur les points suivants pour faciliter l'accès aux connaissances :

- L'utilisation d'icônes du langage VCM permet de faciliter la recherche visuelle de passages traitant d'un sujet donné (par exemple en recherchant le pictogramme du coeur, le médecin peut trouver tous les passages traitant de problèmes cardiaques, ce qui n'est pas possible avec le texte : en effet, "trouble du rythme" fait bien référence à un problème cardiaque, mais ne contient pas la racine "coeur" / "cardi-").
- Il est possible de replier ou de déplier les différentes sections, de manière hiérarchique, en cliquant sur leur titre. Par exemple, sur la figure 6.3, les sections 4.3 (contre-indications) et 4.4 (précautions d'emploi-mises en garde) sont repliées (cela est indiqué par le signe (...)).

6.3.2 "Monsieur VCM"

Le schéma anatomique que nous avons conçu a la forme d'un "bonhomme" stylisé que nous avons appelé "Monsieur VCM". Les différents pictogrammes de localisation et d'étiologie ont été placés sur le schéma (voir figure 6.4). Seuls la tête, les pensées, le corps et un bras du bonhomme sont représentés, délimitant ainsi 5 zones : la tête (avec des localisations comme les yeux, les oreilles, le

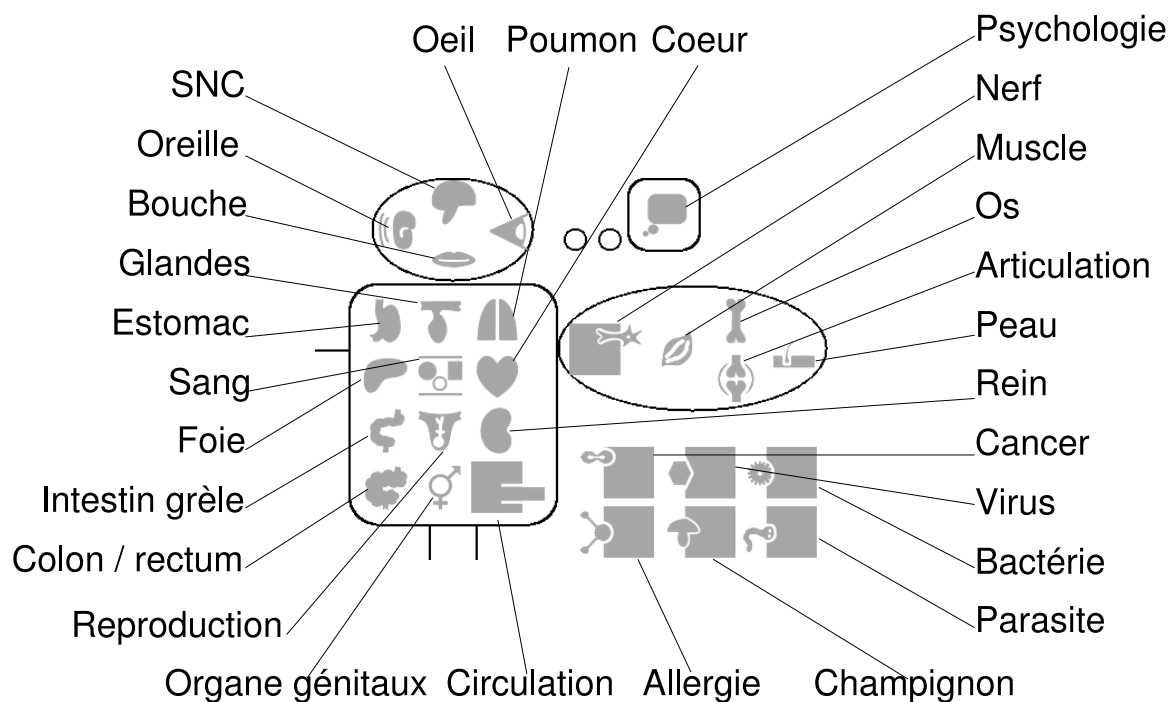


Fig. 6.4 – La répartition des différentes localisations et étiologies sur “Monsieur VCM”.

SNC,...), les pensées (psychiatrie, psychologie, comportement,...), le corps (système digestif, système cardio-pulmonaire, organes sexuels,...), le bras (système nerveux périphérique, os, muscle, peau,...) et un espace en dehors du bonhomme où sont représentées les étiologies (virale, bactérienne,...). En effet les étiologies sont presque toujours des agents extérieurs au patient (à l’exception de l’étiologie tumorale).

Les icônes sont ensuite placées au niveau de la localisation ou de l’étiologie correspondante. Lorsqu’une icône contient à la fois une localisation et une étiologie, elle est présente deux fois. Au contraire, si une pathologie ou un médicament n’a ni localisation ni étiologie, son icône est placée en dessous du bonhomme. “Monsieur VCM” permet de trouver très rapidement les informations correspondant à un système ou une étiologie donné, puisque les mêmes localisations / étiologies sont toujours placées au même endroit. Par exemple, face à un patient asthmatique, le médecin ou le pharmacien regardera en haut à droite du corps du bonhomme pour trouver les contre-indications qui concernent le système respiratoire.

6.3.3 Outil de vérification

L’outil de vérification s’intéresse aux sections contre-indications, interactions médicamenteuses et précautions d’emploi-mises en garde du RCP, l’objectif étant de “faire le tri” le plus vite possible pour éliminer toutes les situations qui ne concernent pas un patient donné. Pour présenter ces sections, nous avons mis en avant le contexte de la prescription : les conditions des contre-indications, interactions médicamenteuses et précautions d’emploi-mises en garde, représentées à l’aide de “Monsieur VCM”. Les précautions d’emploi qui n’ont pas de condition sont regroupées ensemble, sous un contexte “chez tous les patients”, représenté par un rond rouge en VCM.

Les contre-indications, les interactions médicamenteuses et les précautions d’emploi-mises en garde peuvent être représentées chacune sur un “Monsieur VCM” séparé (trois “Monsieur VCM” en tout : un pour les contre-indications, un pour les interactions médicamenteuses et un pour les précautions d’emploi-mises en garde). Mais il est aussi envisageable de les superposer, en utilisant les règles de combinaison d’icônes. Une approche intéressante consiste à superposer sur un même “Monsieur VCM” les contre-indications et les interactions médicamenteuses ; lorsqu’il y a à la fois une contre-indication avec une pathologie et une interaction médicamenteuse avec un médicament portant sur la même localisation, l’icône retenue sera celle de la pathologie. Par exemple s’il y a

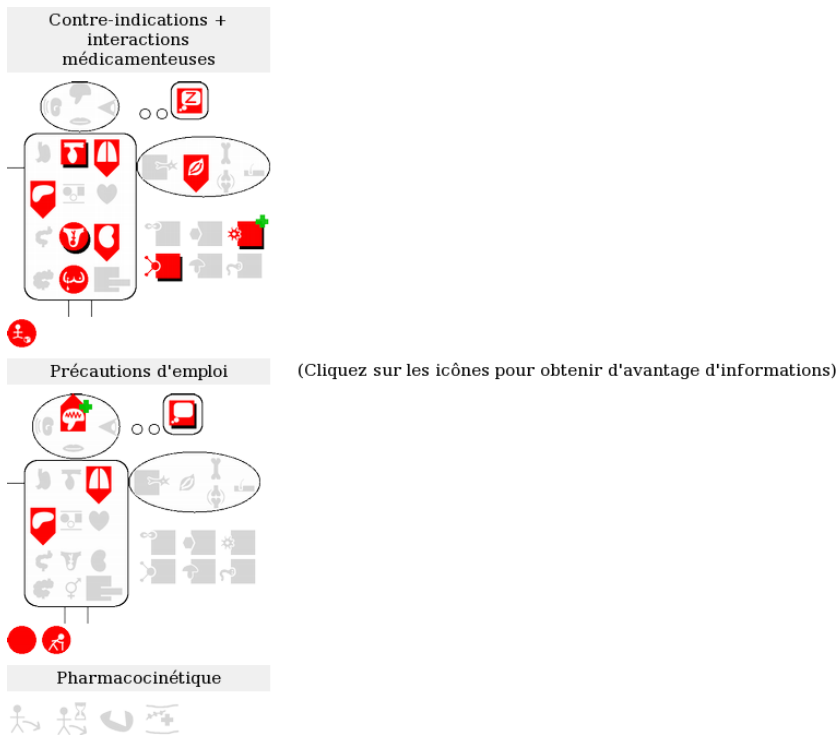


Fig. 6.5 – Copie d'écran de l'outil de vérification, pour le Stilnox (un somnifère).



Fig. 6.6 – Copie d'écran de l'outil de vérification, pour le Stilnox, après avoir cliqué sur l'icône des contre-indications en cas de troubles musculaires.

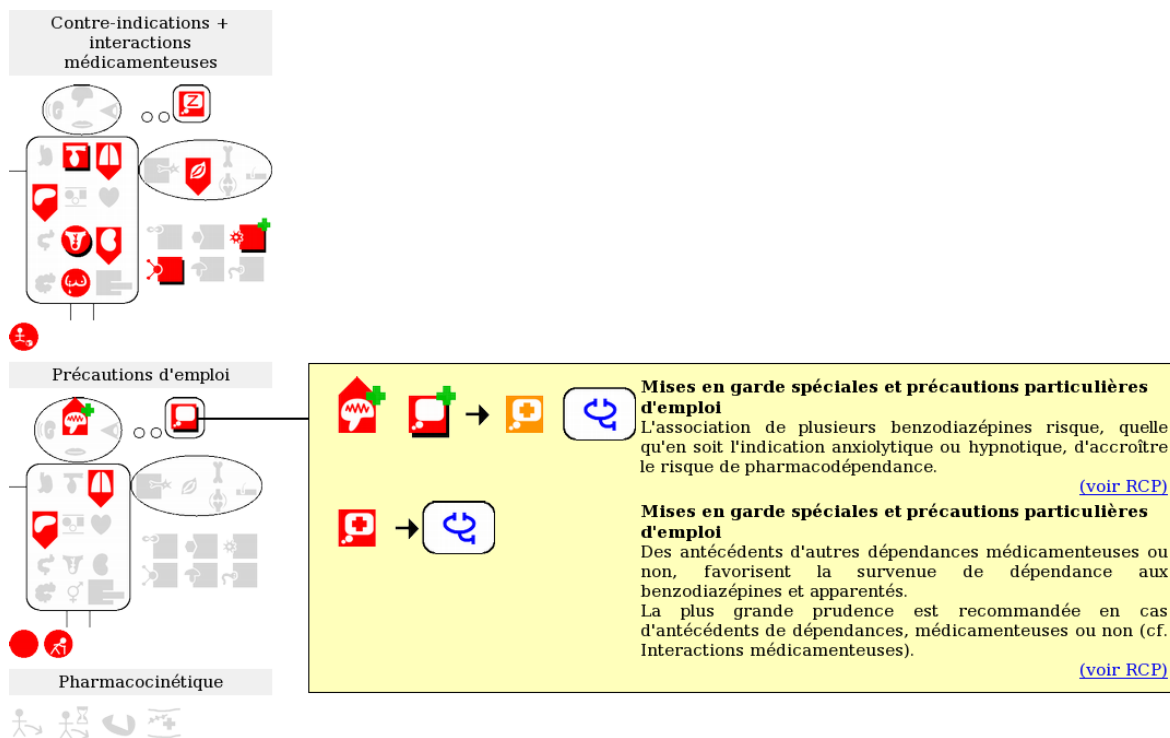


Fig. 6.7 – Copie d’écran de l’outil de vérification, pour le Stilnox, après avoir cliqué sur l’icône des précautions d’emploi en cas de troubles psycho-comportementaux.

contre-indication avec certaines formes de dépression et interactions médicamenteuses avec les ISRS (Inhibiteurs Sélectifs de la Recapture de la Sérotonine, une classe d’antidépresseur), seul l’icône de la dépression figurera sur “Monsieur VCM”. Cependant, si le médecin est face à un patient traité par ISRS, ce patient souffre *a fortiori* de dépression. Le médecin cliquera donc sur l’icône de la dépression pour avoir plus de détails (puisque le patient est concerné), et verra donc bien l’interaction médicamenteuse avec les ISRS.

L’écran de l’outil de vérification se décompose en deux parties (voir figure 6.5). La première partie est placée à gauche de l’écran et comprend plusieurs icônes du langage VCM, organisées en trois blocs : un pour le “Monsieur VCM” des contre-indications et interactions médicamenteuses, un pour le “Monsieur VCM” des précautions d’emploi-mises en garde, et une barre d’icônes de pharmacocinétique. La seconde partie, placée à droite, sera utilisée pour afficher des textes extraits du RCP.

Les icônes de pharmaco-cinétique comprennent, dans l’ordre, le mode d’élimination (rénale, hépatique), la demi-vie du principe actif, l’activité sur les enzymes du cytochrome P450 (activateur ou inhibiteur enzymatique) et la fixation aux protéines plasmatiques. Lorsqu’une spécialité présente une propriété particulière, l’icône correspondante est représentée en vert (convention du langage VCM), et elle est complétée par l’icône du rein ou du foie pour le mode d’élimination, et par une flèche vers le haut ou vers le bas pour l’activité sur les enzymes (vers le haut pour les activateurs, vers le bas pour les inhibiteurs). Lorsqu’il n’y a rien à signaler, l’icône est grisée.

Si l’utilisateur souhaite plus d’information sur une icône, **il suffit de cliquer dessus pour faire apparaître sur la droite une boîte de dialogue contenant le ou les extraits du RCP concernés** (voir figure 6.6), à la fois en VCM et en texte, ainsi qu’un lien “voir RCP” vers le texte intégral du RCP (utilisant l’outil de lecture décrit ci-dessus). Lorsque plusieurs passages du RCP correspondent à l’icône cliquée, les différents passages sont placés les un en-dessous des autres dans la boîte de dialogue (voir figure 6.7).

6.3.4 Outil d’aide à l’éducation du patient

L’outil d’aide à l’éducation du patient a pour objectif de récapituler sur un seul écran tout ce que le médecin ou le pharmacien doit dire à son patient sur un médicament donné. Nous avons choisi

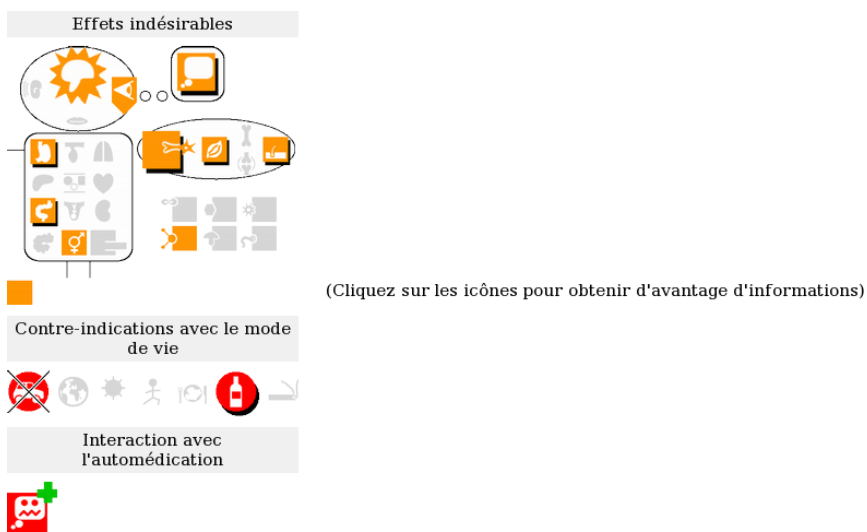


Fig. 6.8 – Copie d'écran de l'outil d'aide à l'éducation du patient, pour le Stilnox (un somnifère).

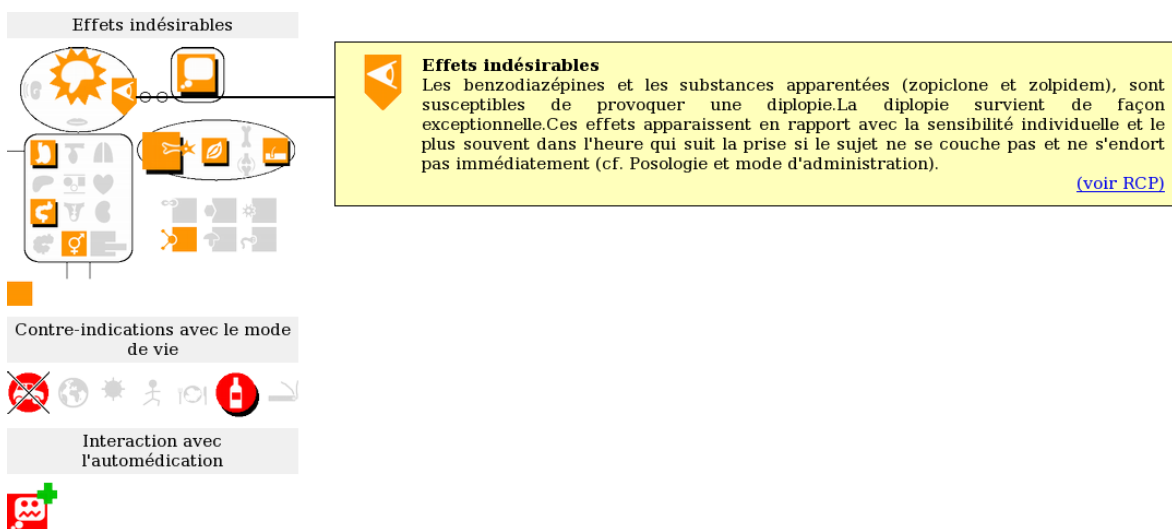


Fig. 6.9 – Copie d'écran de l'outil d'aide à l'éducation du patient pour le Stilnox, après avoir cliqué sur l'icône des effets indésirables oculaires.

de prendre en compte à la fois les connaissances qui répondent à une demande forte de la part du patient, et les connaissances qui sont jugées importantes par le médecin (voir section 3.6.2). Pour notre prototype, nous avons retenu :

- les effets indésirables,
- les modifications à apporter aux habitudes de vie, notamment les risques en cas de conduite automobile, les interactions avec les aliments, l’alcool ou le tabac, les voyages,...
- les interactions médicamenteuses avec des médicaments en vente libre, susceptibles d’être pris en automédication, comme l’aspirine.

Les effets indésirables peuvent être représentés par “Monsieur VCM”. Une difficulté souvent évoquée par les médecins et les pharmaciens est la hiérarchisation des effets indésirables : lesquels sont vraiment importants et doivent être mentionnés au patient et lesquels ne le sont pas ? Les critères suivants peuvent être pris en compte pour déterminer l’importance d’un effet indésirable :

- la fréquence d’apparition,
- la gravité de l’effet,
- l’existence de conduites à tenir permettant de limiter l’effet indésirable (par exemple éviter de s’exposer au soleil) et donnant lieu à un conseil à donner au patient.

Pour faciliter la hiérarchisation des effets indésirables, nous proposons donc de représenter en plus gros les icônes des effets indésirables importants, c’est à dire fréquents, graves ou donnant lieu à un conseil.

L’écran de l’outil d’aide à l’éducation du patient se décompose en deux parties (voir figure 6.8). La première partie est placée à gauche de l’écran et comprend plusieurs icônes du langage VCM, organisées en trois blocs, un pour chaque type d’information : effets indésirables, habitudes de vie et auto-médication. La seconde partie, placée à droite, contiendra les textes extraits du RCP.

Les icônes des habitudes de vie comprennent, dans l’ordre, la conduite automobile, les voyages, l’exposition au soleil, l’activité sportive, l’alimentation, la consommation d’alcool et de tabac. Lorsqu’une habitude de vie pose un problème particulier, son icône est représentée sur un rond rouge (convention du langage VCM) et éventuellement barrée d’une croix (en cas d’interdiction) ; lorsqu’il n’y a rien à signaler, l’icône est grisée. Sur la figure précédente, nous voyons d’un seul coup d’oeil que le Stilnox peut poser des problèmes en cas de conduite automobile ou de consommation d’alcool. Les icônes d’interaction avec l’auto-médication sont les icônes VCM des médicaments correspondants.

Les icônes représentées dans la partie gauche permettent à l’utilisateur de se remémorer les principales propriétés du médicament. Par exemple, il peut sembler évident que le Stilnox, une benzodiazépine utilisée comme somnifère, pose des problèmes en cas de conduite automobile. Cependant la présence d’une icône permet de rappeler au médecin qu’il faut en avertir le patient.

Comme pour l’outil de vérification, si l’utilisateur souhaite plus d’information sur une icône, il suffit de cliquer dessus pour faire apparaître le texte correspondant dans la partie droite de l’écran (voir figure 6.9).

6.3.5 Outil de suivi

L’outil de suivi a pour objectif de récapituler les informations nécessaires au médecin pour le suivi du traitement. Cet outil (voir figure 6.10) se présente de manière similaire aux précédents, avec sur la gauche plusieurs blocs d’icônes, et sur la droite des informations plus précises qui apparaissent lorsque l’on clique sur ces icônes.

Dans notre prototype, cet outil récapitule les effets indésirables ainsi que les risques liés au surdosage (hors effets indésirables), sur 2 “Messieurs VCM”. La comparaison des deux “Messieurs” permet de repérer très facilement les signes propres au surdosage, par exemple pour le Stilnox les signes pulmonaires et vasculaires.



Fig. 6.10 – Copie d’écran de l’outil d’aide au suivi pour le Stilnox.

6.4 Discussion et conclusion

6.4.1 À propos de la méthode

Nous avons proposé une méthode de *Fisheye* avec “Monsieur VCM”. Une autre utilisation du *Fisheye* filtrant pourrait être envisagée : il s’agirait d’afficher les passages du texte concernant un certain profil de patients (personnes âgées, asthmatiques,...), en fonction d’informations entrées par l’utilisateur ou récupérées dans un dossier patient électronique.

La combinaison de plusieurs icônes en une seule est intéressante car elle apporte un gain d’abstraction. Par exemple, en combinant l’icône d’une pathologie cardiaque avec l’icône d’un médicament cardiaque, nous obtenons une icône plus abstraite indiquant un “concept lié au coeur”. Le français ne dispose pas toujours de mots correspondant à ces icônes, cependant nous pensons qu’elles correspondent au mode de pensée des professionnels de santé, en particulier lorsqu’il s’agit d’abstractions de nature anatomique ou étiologique.

6.4.2 À propos des résultats

“Monsieur VCM” a été conçu comme une représentation interactive : en cliquant sur les icônes, le médecin peut accéder aux différents passages du texte. Cependant, puisqu’il rend explicites des connaissances implicites, **“Monsieur VCM” pourrait aussi être utilisé sans interactivité**, par exemple sur papier. Dans ce cas, “Monsieur VCM” permettrait de se rendre très vite compte de l’absence de contre-indication, d’interaction médicamenteuse, de précaution d’emploi-mise en garde ou d’effet indésirable pour une localisation ou une étiologie donnée. En revanche, il serait peu utile dans le cas où une contre-indication, une interaction médicamenteuse, une précaution d’emploi-mise en garde ou un effet indésirable existe, puisqu’il faudrait aller le chercher dans le texte. Il est aussi possible de concevoir un “Monsieur VCM interactif sur papier”, par exemple en numérotant les lignes du texte et en ajoutant à côté des icônes des petits numéros faisant référence aux lignes correspondant dans le texte, à la manière d’un index.

Le prototype que nous avons réalisé implémente plusieurs des outils que nous avons proposés en section 3.7. Nous avons cependant choisi de laisser de côté les outils de navigation, de comparaison et de posologie et plan de prise pour l’instant, pour les raisons déjà évoquées. Par ailleurs, étant pharmacien de formation, j’ai préféré m’intéresser à la validation des ordonnances (outil de vérification) et à l’éducation du patient plutôt qu’au choix du médicament à prescrire (outils de navigation et de comparaison).

L’outil de suivi pourrait être amélioré en ajoutant les précautions d’emploi qui nécessitent un

suivi (surveillance à effectuer régulièrement,...).

En ce qui concerne le support du langage VCM, le prototype est limité à des médicaments qui traitent un état actuel du patient, et à des surveillances qui surveillent la survenue d'un risque. Cependant, il serait intéressant de représenter des médicaments traitant des antécédents, par exemple des antécédents d'infarctus afin d'éviter un nouvel infarctus, ou bien des risques, par exemple des traitements préventifs comme les vaccins. Pour les risques, nous l'avons déjà fait de manière exceptionnelle pour les contraceptifs, que nous avons représentés comme des "traitements du risque de grossesse".

Pour les surveillances, celles qui sont présentes dans les connaissances sur le médicament concernent quasi-exclusivement des surveillances portant sur la survenue d'épisodes iatrogènes, d'où l'utilisation de risques. Cependant, dans le cadre plus large des connaissances médicales, une surveillance peut surveiller l'évolution d'une pathologie, voire d'un antécédent, par exemple suite à un cancer.

6.4.3 Conclusion

Dans ce chapitre, nous avons proposé deux méthodes pour combiner les icônes VCM et le texte : la juxtaposition simple, et "Monsieur VCM", un schéma anatomique interactif. Ces deux méthodes ont ensuite été mises en application lors de la réalisation d'un prototype de base de connaissances sur le médicament, utilisant les données de la base Thériaque. Nous allons maintenant nous intéresser à l'évaluation de VCM et de "Monsieur VCM", dans laquelle nous utiliserons ce prototype.

Chapitre 7

Évaluation du langage VCM

7.1 Introduction et objectifs

L'évaluation du langage VCM et de "Monsieur VCM" a pour objectif :

- de vérifier que le langage VCM peut être appris en quelques heures,
- d'évaluer la compréhensibilité des icônes VCM,
- de comparer la vitesse de lecture des icônes VCM et de "Monsieur VCM" par rapport aux textes utilisés habituellement en clinique.

Pour cela, nous avons réalisé une évaluation dans des conditions contrôlées sur un groupe de médecins généralistes formés à VCM. L'évaluation a eu lieu en trois temps : d'abord une première séance (1/2 journée) pour présenter VCM et "Monsieur VCM", puis un travail personnel des médecins sur 1 mois pour apprendre VCM en utilisant un didacticiel, et enfin une deuxième séance (1 journée) au cours de laquelle nous avons réalisé les évaluations proprement dites.

7.2 Matériel et méthodes

7.2.1 Recrutement des médecins évaluateurs

Nous avons recruté 11 médecins généralistes *via* la SFTG (Société de Formation Thérapeutique du Généraliste, <http://www.unimedia.fr/homepage/sftg/>). Les médecins ont été formés au langage VCM et à "Monsieur VCM", lors de la première demi-journée, puis par un travail personnel sur un didacticiel interactif et un document papier, qui s'est étalé sur un mois. Les médecins ont été indemnisés pour leur participation, grâce à une subvention de la CNAMTS.

7.2.2 Évaluation 1 : compréhensibilité des icônes VCM

L'objectif de cette évaluation est de mesurer la compréhensibilité des icônes VCM. Pour cela, nous avons demandé aux évaluateurs de traduire en français 100 icônes ou phrases VCM, sans limite de temps. Les 100 icônes comprenaient 20 états physiopathologiques et caractéristiques patients, 20 risques, 20 médicaments, 15 surveillances et 25 phrases. Celles-ci ont été choisies de manière à être représentatives de l'ensemble des icônes du langage, contenant à la fois des icônes simples et des icônes plus complexes. L'ordre des icônes était aléatoire, avec un ordre différent pour chaque évaluateur. L'évaluation commençait par 3 icônes d'entraînement (en plus des 100) qui n'ont pas été prises en compte dans les résultats. Cette évaluation s'est déroulée sur papier.

7.2.3 Évaluation 2 : comparaison de la vitesse de lecture avec les icônes VCM par rapport au texte

L'objectif est de comparer la vitesse de lecture des icônes VCM par rapport au texte : lit-on plus vite des connaissances médicales en VCM ou en texte ? en faisant plus ou moins d'erreurs ? Pour

cela, nous avons posé aux évaluateurs des questions précises portant sur des documents médicaux proches des monographies des médicaments, soit en texte, soit en VCM, et nous avons mesuré la justesse des réponses et le temps mis pour répondre. Les documents textes ont été modifiés de sorte à ce que le texte ne contienne pas plus d'informations que les icônes VCM. En effet, nous avons vu que VCM n'est pas aussi précis que le texte; nous avons donc généralisé et raccourci le texte pour obtenir un texte équivalent aux icônes. Par exemple "contre-indiqué en cas d'insuffisance cardiaque congestive" est devenu "contre-indiqué en cas d'insuffisance cardiaque", car VCM ne représente pas l'insuffisance cardiaque congestive. Les textes et les documents VCM étaient représentés "en bloc", sans aucun titre ni classement des différents paragraphes.

7.2.4 Évaluation 3 : comparaison de la vitesse de lecture avec "Monsieur VCM" par rapport au texte

L'objectif est de comparer la vitesse de lecture de "Monsieur VCM" par rapport au texte. La méthodologie est la même que celle de l'évaluation précédente, cependant les documents diffèrent. Les textes n'ont pas été modifiés pour réduire la quantité d'information qu'ils contenaient, et ils ont été découpés en 4 sections : contre-indications, précautions d'emploi-mises en garde, interactions médicamenteuses et effets indésirables, pour arriver à quelque chose de proche de ce que proposent les RCP. Les documents en VCM comprenaient trois "Messieurs VCM" interactifs (un pour les contre-indications et les interactions médicamenteuses, un pour les précautions d'emploi, un pour les effets indésirables). Lorsque le médecin cliquait sur une icône, le texte correspondant apparaissait.

7.2.5 Description des documents et des questions

Les documents utilisés lors des évaluations 2 et 3 étaient des "RCP chimériques", c'est-à-dire des documents proches des RCP mais ne correspondant pas à des médicaments existants, afin d'éviter que les médecins ne fassent appel à leurs connaissances. Les documents étaient composés de paragraphes contenant une seule contre-indication, interaction médicamenteuse ou précaution d'emploi-mise en garde, ou bien un petit nombre de ces éléments, par exemple plusieurs contre-indications cardiaques. Nous avons considéré deux tailles de document : des **documents courts avec 10 paragraphes**, et des **documents long avec 30 paragraphes**.

Les questions correspondaient aux modèles suivants :

- Ce médicament peut-il être prescrit sans précaution particulière chez le patient dans l'état physiopathologique X ?
- Ce médicament peut-il être prescrit sans précaution particulière chez le patient prenant le médicament Y ?
- Ce médicament peut-il provoquer l'effet indésirable Z ?

Les réponses possibles se limitaient à "oui" et "non". Nous avons distingué deux types de question : **les questions à réponse explicite**, pour lesquelles la réponse est donnée de manière explicite dans le texte (par exemple la question était "Ce médicament peut-il être prescrit sans précaution particulière chez le patient asthmatique?" et le document dit "Ce médicament est contre-indiqué en cas d'asthme"), et **les questions à réponse implicite**, pour lesquelles le médecin devait lire la totalité du texte avant de pouvoir déduire la réponse (par exemple, la question est la même que précédemment, le document ne dit rien sur l'asthme, et le médicament peut donc implicitement être prescrit).

7.2.6 Génération des RCP chimériques et des questions

La figure 7.1 montre la méthode utilisée pour générer les RCP chimériques et les questions des évaluations 2 et 3. Tout d'abord, une base de paragraphes a été créée en regroupant les paragraphes présents dans un ensemble de RCP issus de Thériaque et choisis de manière aléatoire. Tous les

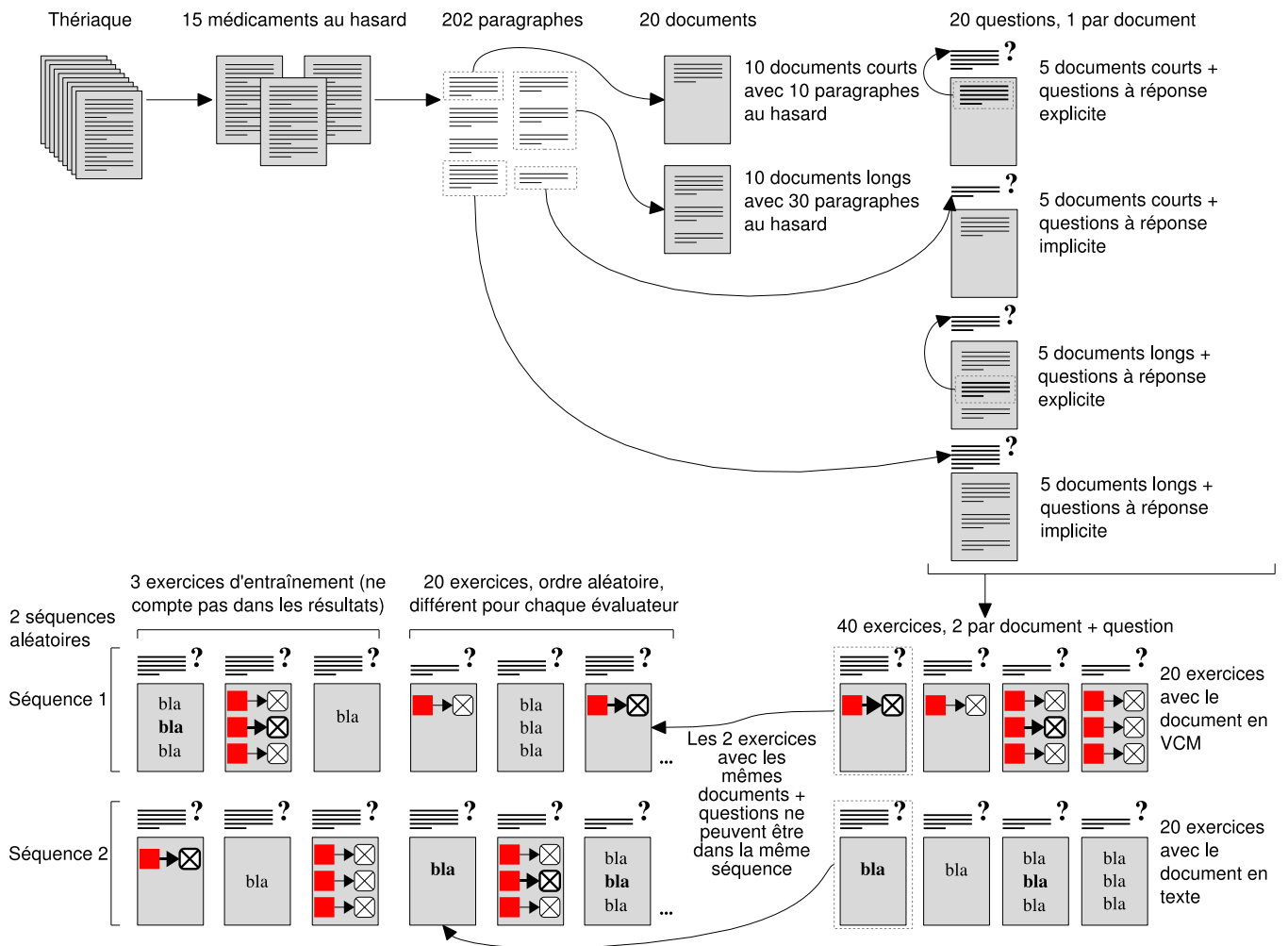


Fig. 7.1 – Méthode utilisée pour générer les RCP chimériques et les questions à partir de la base Thériaque, et pour construire les séquences.

paragraphes exprimant des contre-indications, interactions médicamenteuses, précautions d'emploi-mises en garde ou effets indésirables ont été retenus. Les paragraphes ont été traduits en texte, à l'aide des extraits de textes présents dans Thériaque, et en VCM, à l'aide des données structurées de Thériaque. La traduction en VCM a été réalisée de manière automatique. Les textes ont ensuite été modifiés pour ne pas mentionner le nom du médicament du RCP, lequel a été remplacé par "ce médicament".

Les documents ont ensuite été créés en assemblant 10 ou 30 paragraphes choisis de manière aléatoire dans la base, sans doublon. Chaque document a été revu manuellement, afin de vérifier l'absence d'incompatibilité entre les paragraphes ; par exemple les paragraphes "Contre-indiqué en cas d'asthme" et "En cas d'asthme, diminuer la posologie" ne sont pas compatibles. En cas d'incompatibilité, le second paragraphe a été remplacé par un nouveau paragraphe choisi aléatoirement.

Les questions portaient sur les connaissances contenues dans un seul paragraphe. Pour les questions à réponse explicite, le paragraphe a été choisi aléatoirement parmi ceux du document. Pour les questions à réponse implicite, le paragraphe a été choisi aléatoirement dans la base de paragraphes, en excluant les paragraphes déjà présents dans le document, et en retirant jusqu'à obtenir un paragraphe qui donne lieu à une question à réponse implicite. Lorsque le paragraphe comprenait plusieurs contre-indications, interactions médicamenteuses ou précautions d'emploi-mises en garde, l'une d'elles a été choisie de manière aléatoire.

7.2.7 Schéma des évaluations 2 et 3

Le schéma général est le même pour les évaluations 2 et 3. Chacune comprenait 20 couples (document, question), différents pour les deux évaluations. Ces couples se répartissaient ainsi :

- 5 couples (document court, question à réponse explicite),
- 5 couples (document court, question à réponse implicite),
- 5 couples (document long, question à réponse explicite),
- 5 couples (document long, question à réponse implicite).

Chaque question était posée deux fois, une fois avec le document en VCM et une fois avec le document en texte, dans un ordre aléatoire différent pour chaque évaluateur (voir figure 7.1). Chaque évaluation était découpée en deux séquences de 20 questions, chaque séquence n'incluant qu'une fois chaque document (soit en VCM, soit en texte), en alternant les documents VCM et texte. Trois questions d'entraînement, non prises en compte dans les résultats, ont été ajoutées au début de chaque séquence. Les deux séquences étaient séparées par une pause de 15 minutes, de sorte à ce que les évaluateurs oublient les questions et les réponses. De plus, nous avons dit aux évaluateurs que les questions et documents n'apparaissaient qu'une seule fois chacun, bien que cela soit faux.

Les deux évaluations ont été réalisées sur ordinateur, en utilisant le prototype réalisé au chapitre 6. Les machines utilisées étaient des Macintosh sous MacOSX, de puissances similaires ; les évaluations ne nécessitaient que peu de ressources matérielles. La question était d'abord affichée, et l'évaluateur en prenait connaissance. Puis il appuyait sur un bouton pour faire apparaître le document, cherchait la réponse dans le document et appuyait sur le bouton correspondant à la réponse choisie. L'ordinateur enregistrait alors le temps de réponse, égal à la durée écoulée entre l'affichage du document et la réponse, ainsi que la justesse de la réponse. Ce mode de fonctionnement avait été expliqué aux évaluateurs, et ceux-ci avaient pour ordre de ne pas perdre de temps pendant que le document était visible.

Le critère principal de jugement est le temps de réponse aux questions. Le nombre d'erreurs commises est un critère secondaire.

7.2.8 Analyse statistique des évaluations 2 et 3

L'objectif de l'analyse statistique était de comparer les résultats obtenus avec VCM et avec le texte. Nous avons utilisé un test t apparié pour comparer les temps de réponse. Les 3 facteurs suivants ont été pris en compte : évaluateur, longueur du document, type de question (réponse explicite ou

implicite). L'ANOVA a été utilisée pour mesurer l'effet de ces trois facteurs et de leurs interactions sur le temps de réponse.

Nous avons utilisé un test de Fisher exact pour comparer les taux d'erreur mesurés avec VCM et avec le texte. Ce test a été utilisé à cause du petit nombre de sujets.

Le niveau de signification a été fixé à $\alpha = 0.05$. Les données ont été analysées avec le logiciel R version 2.2.1 [124].

7.3 Résultats

7.3.1 Médecins évaluateurs

L'évaluation s'est déroulée en janvier-février 2006. Nous avons recruté 11 médecins généralistes volontaires de la SFTG, 8 hommes et 3 femmes. La moyenne d'âge est de 51 ans. Ils ont suivi 2 heures de formation lors de la première séance, puis une moyenne de 4 heures (de 2 à 6) de formation sur le didacticiel (d'après les informations recueillies par un questionnaire papier). Ce qui correspond à un total de 6 heures d'apprentissage par médecin en moyenne.

L'acceptation globale du langage VCM par les médecins évaluateurs a été très bonne; les 11 évaluateurs ont été unanimes pour approuver le projet. Ils nous ont aussi fait quelques suggestions intéressantes que nous verrons dans les perspectives (voir section 8.2).

7.3.2 Matériel pédagogique

Le matériel pédagogique que nous avons utilisé pour l'apprentissage de VCM auprès des médecins évaluateurs comprend deux éléments : un manuel papier et un didacticiel (sur CD-ROM). Le manuel d'apprentissage de VCM est reproduit en annexes E, F et G. Il peut aussi être téléchargé, en français ou en anglais, à l'adresse suivante : http://vcm.limbio-paris13.org/guide_vcm/.

Le didacticiel reprend les dictionnaires et le lexique du manuel, et propose un apprentissage de la grammaire et des icônes de VCM en 7 leçons. Chaque leçon comprend une partie "cours" et des exercices interactifs (voir figure 7.2). Enfin, le didacticiel contient plusieurs monographies en VCM.

7.3.3 Évaluation 1 : compréhension des icônes VCM

La figure 7.3 montre un extrait du formulaire papier de l'évaluation de compréhension, et la figure 7.4 les résultats. Les résultats sont hétérogènes, et varient de 59% à 98% de bonnes réponses; cependant nous constatons que seuls quelques évaluateurs ont des mauvais résultats. La moyenne est de 86% de bonnes réponses, et la médiane de 91%.

7.3.4 Caractéristiques des RCP chimériques générées

Pour l'évaluation 2, nous avons extrait 202 paragraphes à partir de 15 RCP, et nous avons construit 20 couples (document, question). Ces 20 documents contenaient un total de 400 paragraphes ($10 \times 10 + 10 \times 30$). Par conséquent, chaque paragraphe était présent en moyenne dans 2 documents différents.

Le tableau 7.1 montre le nombre de mots et d'icônes en moyenne pour les 5 documents de chacune des 4 catégories de longueur et type de question, ainsi que le nombre de mots et d'icônes à lire avant de trouver la réponse, en supposant une lecture linéaire de haut en bas et de droite à gauche. Pour les questions à réponse implicite, il est nécessaire de lire la totalité du document.

En moyenne, nous constatons qu'une icône VCM correspond à 7 mots.

Pour l'évaluation 3, nous avons extrait 200 paragraphes à partir de 14 RCP, et nous avons construit 20 couples (document, question). Comme lors de l'évaluation 2, chaque paragraphe était présent en moyenne dans 2 documents différents.

VCM Lam&Bio

VCM (version 0.9-15-12-2005)

[Introduction \(cliquez ici !\)](#)

Manuel d'apprentissage :

- [Leçon 1 : caractéristiques du patient, pathologies simples et contre-indications](#)
- [Leçon 2 : pathologies complexes](#)
- [Leçon 3 : représentation en "bonhomme"](#)
- [Leçon 4 : antécédents, risques et effets indésirables](#)
- [Leçon 5 : médicaments, concepts liés aux médicaments et interactions médicamenteuses](#)
- [Leçon 6 : surveillances et précautions d'emploi](#)
- [Leçon 7 : vocabulaire et phrases complexes](#)
- [Leçon 8 : conclusion](#)

Manuel de référence :

- [Lexique des pictogrammes](#)
- [Dictionnaire des pathologies](#)
- [Dictionnaire des caractéristiques patient](#)
- [Dictionnaire des traitements](#)
- [Dictionnaire des surveillances](#)

VCM Lam&Bio




Leçon 2 : pathologies complexes

Nous avons vu dans la leçon 1 comment représenter des pathologies ou des symptômes simples. Pour représenter des pathologies (ou symptômes) de façon plus précise, VCM distingue deux types de pathologies : les pathologies génériques et les pathologies spécifiques.

Les pathologies génériques

Les pathologies génériques peuvent s'appliquer à plusieurs localisations anatomico-fonctionnelles, voire à aucune. Par exemple : les insuffisances (insuffisance cardiaque, insuffisance rénale, insuffisance hépatique,...), les infections (infection pulmonaire, infection urinaire,...), les cancers et processus tumoraux (cancer du poumon, cancer de l'intestin,...), les hémorragies (hémorragie cérébrale, hémorragie pulmonaire, hémorragie tout court),...

Pour représenter les pathologies génériques, la forme du carré **rouge** est remplacée par une autre forme qui précise la pathologie. Cela permet d'effectuer de la combinatoire pour réduire le nombre de signes différents. Les formes utilisées pour exprimer des pathologies génériques sont les suivantes :

-  Augmentation / hyper-fonctionnement, par exemple :
-  Hyperthyroïdie (hyper-fonctionnement de la thyroïde)
-  diminution / hypo-fonctionnement / insuffisance, par exemple :

Exercice 1

Que signifient les icônes suivantes ?



(1) Réponse : [Valider](#)
Bonne réponse !



(2) Réponse : [Valider](#)

Fig. 7.2 – Copies d'écran du didacticiel VCM.



A(10) Réponse : _____



BF(11) Réponse : _____

Fig. 7.3 – Extrait de l'évaluation de compréhension.

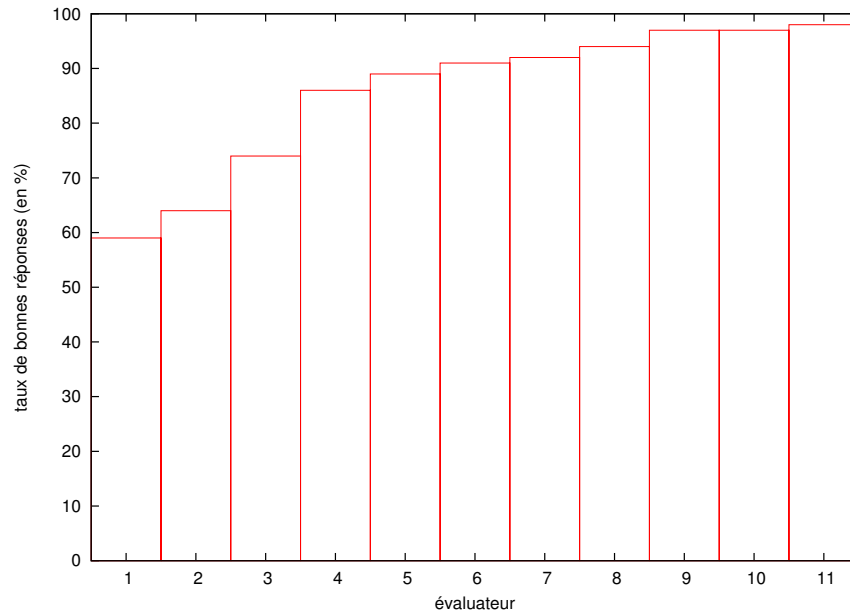


Fig. 7.4 – Résultat de l'évaluation de compréhension.

Les évaluateurs ont été triés par ordre de résultats croissants.

longueur des documents	court		long	
type de question	réponse explicite	réponse implicite	réponse explicite	réponse implicite
nombre de mots moyen	142	153	430	450
nombre de mots moyen avant la réponse	109	153	255	450
nombre d'icônes moyen	23	22	69	67
nombre d'icônes moyen avant la réponse	17	22	40	67

Tab. 7.1 – Caractéristiques moyennes des documents de l'évaluation 2.

longueur des documents	court		long	
type de question	réponse explicite	réponse implicite	réponse explicite	réponse implicite
nombre de mots moyen	438	417	1098	1036
nombre de mots moyen avant la réponse	108	149	243	372
nombre d'icônes moyen	15	16	37	35

Tab. 7.2 – Caractéristiques moyennes des documents de l'évaluation 3.

Vitesse question ZB

Ce médicament peut-il provoquer une tachycardie ?

Oui

Non

Évaluation

Les principaux effets indésirables survenant lors de ce traitement sont généralement des effets indésirables gastro-intestinaux : douleurs de l'abdomen, nausées, vomissements, constipation. L'association de certains traitement nutritionnel ou digestif avec ce médicament impose de revoir le plan de prise.

Ce médicament doit être utilisé avec prudence en cas d'insuffisance hépatique, en raison d'un risque d'accumulation du principe actif.

L'administration de ce médicament peut entraîner une tachycardie.

Cette spécialité ne doit être prescrite au cours de la grossesse qu'en cas de nécessité absolue.

Ce médicament est susceptible de produire des effets indésirables neurologiques, en particulier de la somnolence.

Certaines infections virales peuvent exceptionnellement être à l'origine de graves complications infectieuses cutanées avec ce médicament. Ce médicament ne doit pas être prescrit de ce cas.

Ce médicament ne doit jamais être prescrit en cas de troubles uréthro-prostatiques.

Avec ce médicament, des réactions cutanées phototoxiques et/ou photosensibilisantes sont observés.

Ce médicament n'a pas été évalué chez le nourrisson.

Vitesse question ZA

Ce médicament peut-il provoquer une tachycardie ?

Oui

Non

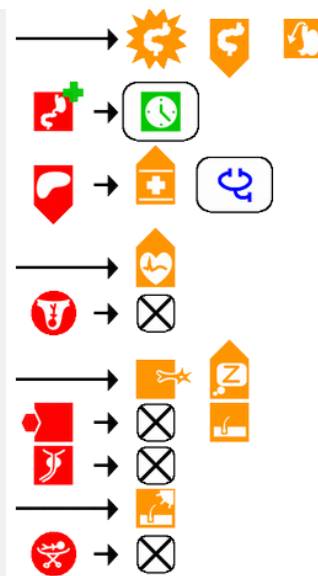


Fig. 7.5 – Copies d'écran de l'évaluation de vitesse.

À gauche avec le document en texte, à droite avec le document en VCM.

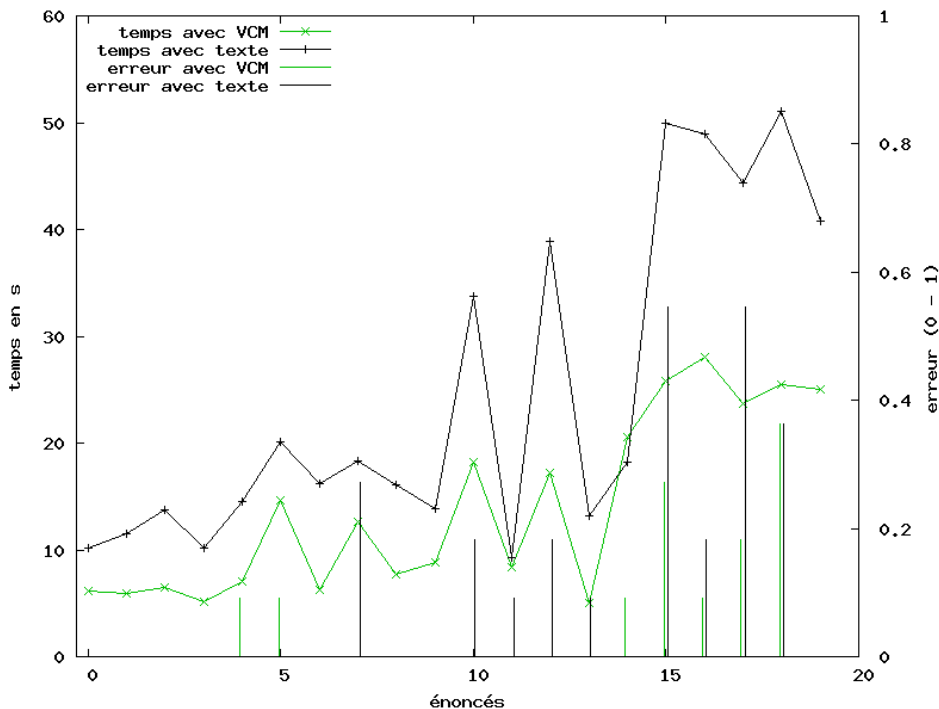


Fig. 7.6 – Résultats de l'évaluation de la vitesse de lecture de VCM.

Les 20 questions figurent en abscisse ; les courbes indiquent le temps mis pour répondre pour chaque question et les barres verticales indiquent le taux d'erreur. Les 10 premières questions portent sur un document court, les 10 dernières sur un document long. Les questions 0 à 4 et 10 à 14 ont leur réponse explicitement mentionnées dans le document, et pour les question 5 à 9 et 15 à 19, la réponse est implicite.

Bonhomme démo 2
Ce médicament peut-il provoquer des hépatites ?

Oui

Non

Évaluation
(titres - passages importants - texte intégral)

1 Contre-indications
Il n'existe pas de données fiables de tératogénèse chez l'animal. Il n'existe pas actuellement de données suffisamment pertinentes pour évaluer un éventuel effet malformatif ou foetotoxique de ce médicament lorsqu'il est administré pendant la grossesse. En conséquence, l'utilisation de ce médicament est généralement déconseillée pendant la grossesse.

2 Mises en garde spéciales et précautions particulières d'emploi
En raison du risque potentiel de complications thromboemboliques, une surveillance destinée à détecter les signes précoces de coagulopathie thrombotique ou de coagulation de consommation doit être initiée à l'aide de tests biologiques appropriés lors de l'administration de ce médicament chez des patients en période postopératoire. Dans cette situation, le bénéfice du traitement par cette spécialité doit être évalué par rapport au risque de ces complications. Il n'est pas recommandé d'injecter ce médicament en sous-cutané dans les zones de peaux infectées ou inflammatoires. Ce médicament doit être utilisé avec prudence chez les malades ayant des antécédents de convulsions ou des facteurs prédisposant à la survenue de convulsions.

3 Interactions avec d'autres médicaments et autres formes d'interactions
L'administration de ce médicament n'est pas recommandée chez les malades sous anticoagulants. L'association de ce médicament et des antivitamines K, se traduisant par une majoration de l'effet anticoagulant, demande de prendre certaines précautions d'emploi. Il existe un risque hémorragique. Il convient d'effectuer des contrôles plus fréquents du taux de prothrombine et de l'INR et, si besoin, d'adapter la posologie de l'anticoagulant.

4 Effets indésirables
L'administration de ce médicament, en particulier si la solution est hypertonique, peut provoquer une douleur locale, une veinite, voire une thrombophlébite ou une nécrose en cas

Bonhomme démo 1
Ce médicament peut-il provoquer des hépatites ?

Oui

Non

Contre-indications + interactions médicamenteuses

Précautions d'emploi

Effets indésirables

Contre-indications avec le mode de vie

Effets indésirables
Des cas d'hépatites, parfois fatales, ont été observés au cours de traitement par ce médicament. [\(voir RCP\)](#)

Fig. 7.7 – Copies d'écran de l'évaluation de "Monsieur VCM".

À gauche avec le document en texte, à droite avec le document utilisant "Monsieur VCM".

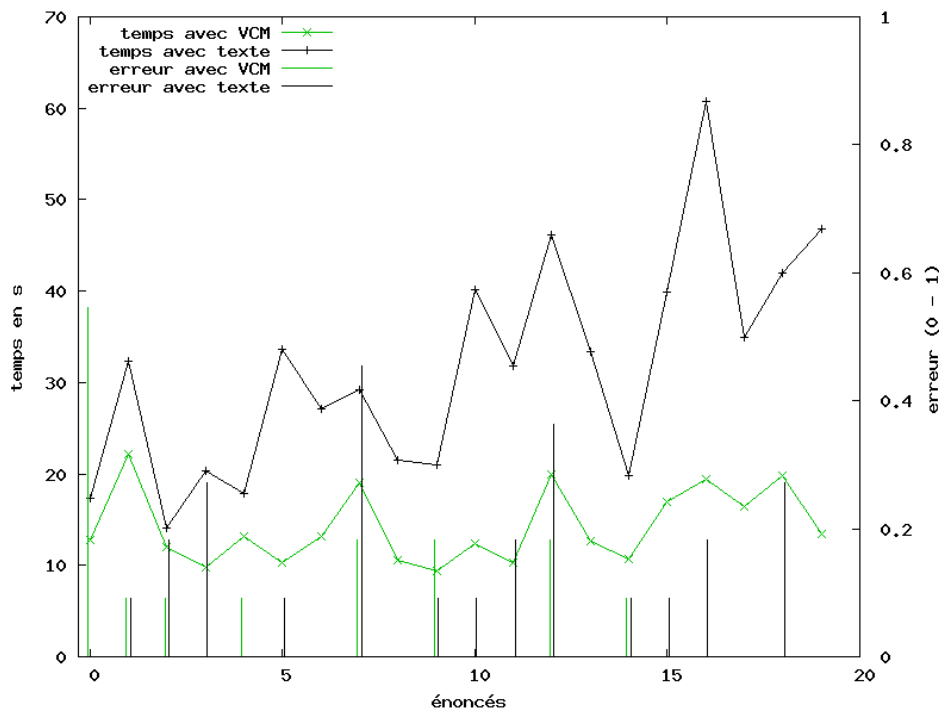


Fig. 7.8 – Résultats de l'évaluation de "Monsieur VCM".

Les 20 questions figurent en abscisse ; les courbes indiquent le temps mis pour répondre pour chaque question et les barres verticales indiquent le taux d'erreur. Les 10 premières questions portent sur un document court, les 10 dernières sur un document long. Les questions 0 à 4 et 10 à 14 ont leur réponse explicitement mentionnées dans le document, et pour les question 5 à 9 et 15 à 19, la réponse est implicite.

Le tableau 7.2 montre le nombre de mots et d'icônes en moyenne pour les 5 documents de chacune des 4 catégories de longueur et type de question, ainsi que le nombre de mots à lire avant de trouver la réponse, en supposant une lecture linéaire de haut en bas et de droite à gauche, limitée à la ou les sections du document concernées par la question (par exemple la section contre-indications n'est pas concernée par une question portant sur les effets indésirables).

Les documents courts contiennent en moyenne 15 icônes sur les "Messieurs VCM", et les longs 36 ; ce chiffre est inférieur à 45 (= 3 x 15) car certaines icônes ont été combinées entre elles. En moyenne, les "Messieurs VCM" contenaient 1 icône pour 30 mots.

7.3.5 Évaluation 2 : comparaison de la vitesse de lecture avec les icônes VCM par rapport au texte

La figure 7.5 montre des copies d'écran de l'évaluation de vitesse. Les résultats montrent qu'en moyenne les médecins vont 1,8 fois plus vite pour répondre à une question ($p < 10^{-16}$, $N = 11$, test t apparié) et font 2,1 fois moins d'erreurs ($p = 0.0025$, $N = 11$, test de Fisher exact) avec VCM par rapport à un texte brut véhiculant les mêmes connaissances. On constate sur la courbe des résultats (figure 7.6) que le temps de réponse et le nombre d'erreurs varient beaucoup en fonction de la longueur du document et du type de question (réponse explicite ou implicite dans le document). L'analyse statistique montre que ces deux facteurs ont une forte influence ; par contre il n'y a pas d'interaction entre eux. Les variations sont similaires avec le texte et avec VCM ; le temps de réponse avec VCM semble proportionnel au temps de réponse avec le texte.

7.3.6 Évaluation 3 : comparaison de la vitesse de lecture avec "Monsieur VCM" par rapport au texte

La figure 7.7 montre des copies d'écran de l'évaluation de "Monsieur VCM". Les résultats montrent qu'en moyenne les médecins vont 2,2 fois plus vite ($p < 10^{-16}$, $N = 11$, test t apparié) et font 1,7 fois moins d'erreurs ($p = 0.03398$, $N = 11$, test de Fisher exact) avec "Monsieur VCM" par rapport à un texte organisé en 4 sections.

On constate sur la courbe des résultats (figure 7.8) qu'avec le texte, le temps de réponse varie beaucoup en fonction de la longueur du document et du type de question (réponse explicite ou implicite dans le document). En revanche, avec VCM, le temps de réponse reste quasi-constant quelle que soit la longueur du document, et quel que soit le type de question.

L'analyse statistique confirme cela : les facteurs "longueur des documents" et "type de question" ont une forte influence sur la lecture du texte. Par contre, le facteur "longueur du document" n'a qu'une faible influence sur VCM, et le facteur "type de question" n'a pas d'influence.

7.4 Discussion et conclusion

7.4.1 À propos de la méthode

Nous avons évalué VCM et "Monsieur VCM" dans des conditions contrôlées, qui diffèrent des conditions cliniques, cependant cela était nécessaire afin d'avoir une comparaison équitable entre VCM et le texte. Nous avons trouvé peu d'évaluations d'approches graphiques en médecine qui se placent dans des situations cliniques. Et celles-ci se limitent souvent à des questionnaires de satisfaction ; c'est le cas de Wachter *et al.* [157] pour l'évaluation d'une approche graphique pour afficher des données patients en pneumologie, ou de Burkhard [21] pour l'évaluation d'une représentation des différentes étapes d'un projet médical.

Les documents et les questions ont été générés de manière totalement aléatoire. Pour les générer, nous avons utilisé les connaissances structurées de la base Thériaque, qui sont indépendantes à la fois du texte et de VCM. Les documents VCM et textuels ont été générés automatiquement à partir

de ces connaissances structurées (lesquelles incluent des textes extraits du RCP). **L'utilisation des RCP chimériques présente deux avantages pour notre évaluation** : d'une part les médecins ne peuvent pas faire appel à leurs connaissances du domaine, et ils sont donc obligés de chercher la réponse dans le document, d'autre part comme il s'agit de documents créés de toutes pièces, il est possible de fixer certains paramètres, comme nous l'avons fait pour la longueur du document.

Dans la littérature, la plupart des études n'utilisent pas des documents aléatoires, mais des documents choisis par les concepteurs de l'étude. Cependant, l'impact du choix des documents reste inconnu : les concepteurs risquent de choisir des documents qui sont favorables à leur représentation graphique, et il est difficile de savoir comment le langage graphique se comporterait sur d'autres documents.

Elting *et al.* [52] ont proposé une méthode portant elle aussi sur des données chimériques : pour mesurer l'impact du mode de présentation graphique sur la décision d'interrompre ou non un essai clinique avant son terme, les auteurs construisent un "essai clinique hypothétique", à rapprocher des RCP chimériques. Cependant, nous n'avons pas trouvé dans la littérature d'autres évaluations mettant en jeu des connaissances chimériques, lesquelles sont plus difficiles à construire que des données chimériques.

La plupart des évaluations de langages graphiques médicaux se sont intéressées au patient [73, 65, 141], et ont comparé le texte avec le même texte agrémenté d'icônes. Le critère principal étaient la compréhension ou l'observance aux recommandations. Le même type d'étude a été utilisé par B. Preiss *et al.* pour évaluer UVAL-MED sur un groupe d'étudiants en médecine [116]. Les critères principaux étaient la compréhension et le rappel.

Cependant, parce qu'elles comparent le texte seul avec le texte plus le langage graphique, ces évaluations ne prennent pas en compte les difficultés éventuellement rencontrées pour comprendre le langage graphique : si l'évaluateur ne comprend pas une icône, il lira le texte correspondant et aucune erreur ne sera rapportée. Cela serait un biais important dans l'évaluation d'un langage nécessitant une phase d'apprentissage comme VCM.

La plupart des études n'utilisent pas des documents aléatoires, mais des documents choisis par les concepteurs de l'étude. Cependant, l'impact du choix des documents reste inconnu : les concepteurs risquent de choisir des documents qui sont favorables à leur représentation graphique, et il est difficile de savoir comment le langage graphique se comporterait sur d'autres documents.

Lorsque nous avons mis au point l'étude, nous n'avions aucune idée de la variance que nous allions obtenir. C'est pourquoi il n'était pas possible de calculer le nombre de sujet *a priori*, et celui-ci a été déterminé en fonction des aspects pratiques et des contraintes budgétaires. Cependant, avec seulement 11 évaluateurs et 2x20 questions (20 questions posées deux fois chacune), les différences observées sont fortement significatives. Dans la littérature, nous trouvons des études avec un nombre d'évaluateurs de cet ordre de grandeur, par exemple 8 évaluateurs et 2x16 questions [32], 34 évaluateurs et 4x1 questions [52].

7.4.2 À propos des résultats

7.4.2.1 Temps d'apprentissage

Le temps d'apprentissage que nous avons mesuré, 6h en moyenne, est rapide, d'autant plus que l'apprentissage a été "dilué" sur un mois. Le temps laissé pour l'apprentissage, un mois, a été jugé trop long : les médecins, enthousiastes, ont dans l'ensemble très vite commencé leur apprentissage mais avaient oublié une partie de ce qu'ils avaient appris le jour de l'évaluation.

7.4.2.2 Compréhensibilité des icônes

Les médecins ont en moyenne compris 86% des icônes. On constate qu'un petit nombre d'évaluateurs tirent la moyenne vers le bas avec des scores inférieurs à 80% de bonnes réponses ; ceux-ci semblent avoir insuffisamment travaillé l'apprentissage du langage VCM (l'un d'eux n'y a passé que 2

heures, et un autre reconnaît avoir un mauvais niveau de connaissance de VCM lors du questionnaire d'entrée).

Les résultats de l'évaluation de compréhensibilité nous permettront de voir quelles icônes posent problème et doivent être revues. Par contre, il est difficile d'interpréter le résultat global faute de pouvoir le comparer à une référence.

De plus, il paraît surprenant que les médecins fassent moins d'erreur lors des évaluations 2 et 3 avec VCM qu'avec le texte, alors que leur compréhension du langage VCM semble imparfaite, puisqu'un médecin a 14% (= 100 - 86) de chance de se tromper lorsqu'il lit une icône VCM! Ce résultat paradoxal peut s'expliquer de plusieurs manières :

- L'évaluation 1 a eu lieu en premier, et a pu mettre aux médecins de se "rafraîchir" la mémoire, ou de terminer leur apprentissage de VCM pour ceux qui aurait insuffisamment travaillé sur le didacticiel.
- Le langage VCM n'a pas été conçu pour être traduit en français, mais pour être "compris" des médecins. En regardant les réponses des médecins, nous avons l'impression qu'ils ont parfois eu des difficultés à exprimer par des mots ce qu'ils avaient compris. Par exemple, l'un d'eux a répondu "danger exposition soleil" pour "risque de photosensibilisation".
- Une compréhension totale du langage VCM n'est peut-être pas nécessaire pour que celui-ci puisse aider le médecin pour répondre à des questions cliniques.

7.4.2.3 Temps de réponse

L'objectif originel de VCM était d'accélérer l'accès aux connaissances médicales par les professionnels de santé. Au vu des résultats, cet objectif apparaît atteint : **les icônes VCM sont lues en moyenne 1,8 fois plus rapidement qu'un texte équivalent, et "Monsieur VCM" permet de répondre à des questions cliniques 2,2 fois plus rapidement qu'à partir d'un texte.**

L'évaluation 3 montre que **le facteur "type de question" n'a pas d'influence sur "Monsieur VCM"**, alors que, avec le texte, répondre aux questions à réponse implicite nécessite significativement plus de temps. Cela confirme notre hypothèse sur l'importance de la représentation explicite des connaissances implicites (voir section 5.2) : si le facteur "type de question" n'a pas d'influence sur "Monsieur VCM", c'est parce que "Monsieur VCM" représente de manière explicite l'absence de contre-indication, interaction médicamenteuse, précaution d'emploi-mise en garde ou effet indésirable concernant une localisation ou une étiologie donnée (voir section 6.2.3). Par conséquent, vis-à-vis de "Monsieur VCM", toutes les questions sont à réponse explicite.

Le facteur "longueur du document" n'a que peu d'influence sur "Monsieur VCM", car la taille du schéma est fixe et n'augmente pas avec la longueur du document. Cependant, un document plus long conduit à un nombre plus important d'icônes sur le schéma (voir tableau 7.2), d'où l'influence de la longueur.

7.4.2.4 Taux d'erreur

Comme les médecins connaissent moins bien le langage VCM que les textes médicaux, nous craignons qu'ils fassent plus d'erreurs avec VCM. Les résultats montrent l'inverse : **les médecins ont fait entre 1,7 et 2,1 fois moins d'erreur avec VCM.** Cette réduction du nombre d'erreurs peut s'expliquer par les caractéristiques propres du langage VCM, et notamment sa capacité à représenter de manière explicite une partie des connaissances qui sont implicites dans les textes. Elle peut aussi s'expliquer par le gain de temps : si, comme nous l'avons vu dans la littérature, le médecin alloue une durée fixe, ne dépassant pas 2 minutes, pour sa recherche [53], ce délai a moins de chance d'être dépassé avec VCM qu'avec le texte.

Les taux d'erreur obtenus lors des évaluations (9% lors de l'évaluation 2 et 10% lors de l'évaluation 3) sont élevés : le médecin ne se trompe pas une fois sur dix lorsqu'il consulte une base de connaissance! Ce taux élevé peut s'expliquer par les raisons suivantes :

- Il s'agissait d'une évaluation, sans patient, donc les erreurs n'avaient pas de conséquences cliniques. Cela a pu inciter les médecins à être moins précautionneux.

- Le nombre de questions posé (92 en tout) est important, ce qui a pu fatiguer les médecins, et diminuer leur attention.
- Les documents étant des RCP chimériques, les médecins ne pouvaient pas s'appuyer sur leurs connaissances. Or, lorsqu'un médecin consulte une base de connaissances, c'est souvent pour vérifier une information qu'il connaît déjà partiellement : la question que le médecin se pose n'est pas toujours "Puis-je prescrire ce médicament chez le patient X ?" mais plutôt "Je crois que ce médicament est (ou n'est pas) contre-indiqué chez le patient X, je vais vérifier". Dans ce cas, le médecin consacrera plus ou moins de temps à sa recherche selon la réponse attendue : s'il pense qu'il y a contre-indication mais qu'il ne trouve rien dans le texte, il regardera plus longtemps.

Lors de l'évaluation 3, le nombre d'erreurs avec VCM est un peu plus important que lors de l'évaluation 2, ce qui se traduit par une diminution du rapport du nombre d'erreurs texte / VCM (1,7 au lieu de 2,1). Ce nombre d'erreurs plus important s'explique par la première question, sur laquelle nous avons un grand nombre d'erreurs en VCM (voir figure 7.8). Cette question était un peu particulière car elle portait sur les risques encourus en cas de surdosage, et était placée à un endroit particulier sur "Monsieur VCM" des précautions d'emploi (et pas celui des effets indésirables), d'où un nombre d'erreurs important.

7.4.2.5 Classification des erreurs

J. Zhang *et al.* ont proposé une taxonomie des erreurs médicales. Ils considèrent deux types d'erreur : **les lapsus** (exécution incorrecte d'une séquence d'actions correcte, *slip*) et **les fautes** (exécution correcte d'une séquence d'actions incorrecte, *mistake*), à chacun des 7 niveaux de la théorie de l'action de Norman : objectif, intention, spécification de l'action, exécution, perception, interprétation et évaluation [164].

Dans les évaluations 2 et 3, 4 causes d'erreur peuvent être distinguées. Pour les questions à réponses explicites, l'évaluateur peut ne pas avoir vu la réponse dans le document (**lapsus de perception** dans la taxonomie de Zhang). Pour les documents en VCM, l'évaluateur peut ne pas avoir compris la signification d'une icône (**faute d'interprétation**). Lorsque la recherche est longue, notamment pour les documents longs, les documents textes et les questions à réponse implicite, l'évaluateur peut laisser tomber la lecture du document avant d'avoir trouvé la réponse, et changer son intention de 'rechercher la réponse dans le document' à 'répondre en utilisant des heuristiques, ses connaissances personnelles ou son intuition' (**faute d'intention**). Enfin, l'évaluateur peut ne pas interpréter le document de la même manière que je l'ai interprété lors de la conception de la base d'évaluation ; par exemple, un médecin a considéré qu'un médicament pouvant causer des troubles rénaux était contre-indiqué en cas d'insuffisance rénale, bien que ce ne soit pas précisé (**faute d'évaluation**).

Ces fautes d'évaluation ont été enregistrées comme des erreurs, bien qu'il s'agisse en fait d'interprétations différentes. Cependant, comme les documents VCM et texte sont équivalents, les évaluateurs auraient dû les interpréter de la même manière, et donc faire autant de fautes d'évaluation avec le texte qu'avec VCM.

Étant plus rapide à lire et moins volumineux, un langage graphique comme VCM peut réduire le risque de lapsus de perception et de faute d'intention. Cependant, le risque de faute d'interprétation est plus important, en particulier si l'évaluateur connaît mal le langage graphique. Nos résultats montrent une réduction du risque global d'erreur, ce qui va dans le sens de l'opinion émise par J. Zhang *et al.* : "la solution aux erreurs médicales n'est pas médicale mais cognitive" [163].

7.4.3 Conclusion

Nous avons évalué VCM et "Monsieur VCM" dans des conditions contrôlées, sur un petit groupe de médecins généralistes ayant reçu 6h de formation. Les résultats montrent que le langage VCM permet aux professionnels de santé de répondre à des questions cliniques environ deux fois plus vite qu'avec les textes médicaux utilisés d'ordinaire, et en faisant deux fois moins d'erreurs. Une prochaine étape

pourrait être l'évaluation de VCM et "Monsieur VCM" sur un nombre plus important de sujets, ou bien la validation en situation clinique. Cependant cela dépasse le cadre de cette thèse.

Chapitre 8

Discussion et conclusion

Les méthodes utilisées et les résultats obtenus ont déjà été discutés dans chacun des 4 chapitres précédents. Dans ce chapitre, nous comparerons tout d'abord VCM aux jeux de pictogrammes ou langages graphiques médicaux existants. Tout au long de cette thèse, nous nous sommes placés dans un cadre multi-disciplinaire. Nous verrons donc successivement les perspectives dans chacune des trois disciplines : (a) en médecine, avec l'extension du langage VCM à l'ensemble de la médecine et aux différents professionnels de santé, (b) en sciences cognitives, avec la mise au point d'une méthode rigoureuse pour concevoir des langages graphiques et l'évolution vers un langage culturellement indépendant, et (c) en informatique, avec la traduction automatique des textes en VCM et l'utilisation des langages graphiques pour la représentation des ontologies.

8.1 Comparaison avec les jeux de pictogrammes ou langages graphiques médicaux existants

Dans l'état de l'art, nous avons vu qu'il existait plusieurs jeux de pictogrammes et une version préliminaire d'un langage graphique médical. **Les pictogrammes de l'USP** (voir section 2.2.4.5) visent un public différent de VCM : ils s'adressent au patient et non au professionnel. Ils se limitent donc à la posologie, la conservation du médicament et à la manière de le prendre, ainsi qu'à quelques contre-indications ou effets indésirables simples à même d'être appréhendés par le patient (par exemple, grossesse ou somnolence). L'objectif n'est pas d'accélérer l'accès aux connaissances comme le fait VCM, mais d'attirer l'attention du patient et de faciliter la compréhension, dans le but d'améliorer l'observance.

Stabilis 2 (voir section 2.2.4.4) propose, comme VCM, des pictogrammes pour représenter les classes thérapeutiques, cependant Stabilis 2 ne prend pas en compte les aspects cognitifs sous-jacents à la visualisation. En particulier, les pictogrammes multiplient les jeux de mots (par exemple un crabe pour représenter les cancers), et le style graphique choisi, pictogrammes en couleur et assez détaillés, rend difficile leur utilisation en grand nombre.

Ni les pictogrammes de l'USP ni Stabilis 2 ne disposent de grammaire : il s'agit d'ensembles de pictogrammes sans réelle possibilité d'extension, alors que la grammaire de VCM permet d'étendre facilement le langage. Par exemple, il est possible de rajouter un pictogramme dans VCM, puis d'utiliser la grammaire du langage pour générer plusieurs nouvelles icônes en combinant ce nouveau pictogramme à ceux existants.

UVAL-MED (voir section 2.2.4.6) propose une grammaire graphique qui reste limitée. Cependant, contrairement à VCM, l'objectif d'UVAL-MED n'est pas d'accélérer la lecture mais plutôt de faciliter l'apprentissage des connaissances. Le langage est limité aux concepts de maladie et de symptôme, tandis que VCM peut aussi représenter des traitements médicamenteux, des surveillances et des actions du médecin. Enfin, bien que prometteur en 1994, UVAL-MED n'a pas donné lieu à des développements ultérieurs.

8.2 Perspectives médicales

8.2.1 Amélioration et extension du langage VCM

Le langage VCM est certainement perfectible ; quelques suggestions recueillies lors de l'évaluation auprès des médecins nous seront d'une grande utilité pour l'amélioration du langage. Par ailleurs, certaines parties des connaissances sur le médicament sont actuellement insuffisamment couvertes par le langage VCM ; c'est le cas des posologies et des plans de prise, ou des niveaux de contre-indication (contre-indication absolue, relative,...) par exemple.

Pour l'instant, nous nous sommes focalisé sur les connaissances sur le médicament. Celles-ci font appel aux principaux concepts médicaux, comme les pathologies, les risques ou les médicaments, que l'on retrouve d'une manière plus générale dans les connaissances médicales. Cependant, pour étendre VCM à l'ensemble des connaissances médicales, et plus seulement aux connaissances sur le médicament, le langage doit être étendu. Par exemple, les actes chirurgicaux devraient être plus détaillés et peut-être disposer d'une couleur spécifique.

Il est aussi possible de concevoir des versions spécialisées de VCM, par exemple pour une spécialité médicale (cardiologie,...).

8.2.2 Utilisation de VCM dans les logiciels et outils médicaux

Le langage VCM pourrait être utilisé dans d'autres logiciels médicaux que les bases de connaissances sur le médicament. Les sections 2.3.8 et 3.7 proposent plusieurs suggestions pour des outils de visualisation des connaissances médicales ; certaines pourraient s'appuyer sur le langage VCM ou "Monsieur VCM". Nous allons détailler quelques exemples.

Les guides de bonnes pratiques ou les logiciels intégrant ces guides pourrait utiliser VCM, afin d'aider le médecin à trouver le passage du guide qui concerne son patient, sans pour autant lui donner l'impression que l'ordinateur "réfléchit" à sa place.

Un logiciel de **dossier patient informatisé** pourrait représenter sur un "bonhomme" les pathologies et les traitements du patient, pour en avoir une vision d'ensemble. De plus, cela faciliterait la détection des contre-indications et interactions médicamenteuses : il suffirait alors de comparer le "bonhomme" du dossier patient avec celui des contre-indications et interactions médicamenteuses d'un médicament.

Les outils de recherche de documents médicaux (comme le moteur CISMef, <http://www.chu-rouen.fr/cismef/>) pourraient utiliser des icônes VCM pour caractériser les documents retrouvés. Nous avons proposé "Monsieur VCM" pour indexer les passages d'un texte médical en fonction des localisations anatomico-fonctionnelles et des étiologies (voir section 6.3.2) ; de la même manière "Monsieur VCM" pourrait indexer un ensemble de documents et permettre une recherche graphique en fonction de l'anatomie et / ou de l'étiologie.

Un outil de comparaison permettant de comparer plusieurs médicaments ou classes thérapeutiques proches pourrait utiliser des "Messieurs VCM".

VCM et "Monsieur VCM" permettent aussi de "**fusionner**" des connaissances médicales, ce qui n'est pas possible avec le texte : par exemple, "Monsieur VCM" fusionne "contre-indiqué en cas d'insuffisance cardiaque" et "contre-indiqué en cas de troubles du rythme" sous une icône signifiant "contre-indiqué dans plusieurs pathologies cardiaques". Il serait ainsi possible de fusionner les propriétés de plusieurs médicaments, pour représenter les contre-indications d'une classe thérapeutique, ou bien de l'ensemble de l'ordonnance du patient.

8.2.3 Utilisation de VCM sur d'autres supports

Dans notre prototype, nous avons utilisé le langage VCM sur un support informatique standard, cependant d'autres supports sont envisageables.

Sur papier, les utilisations sont plus limitées à cause de l'interactivité très réduite, cependant le papier possède un avantage sur l'écran : la résolution est très supérieure. Il est donc possible

d'imprimer des icônes VCM ou des "Messieurs VCM" en beaucoup plus petit qu'à l'écran, tout en restant lisible. Une surface de l'ordre de quelques centimètres carrés apparaît suffisante pour un "Monsieur VCM", ce qui rend son utilisation possible dans des livres comme le Vidal (nous avons vu à la section 6.4.2 que "Monsieur VCM" gardait un intérêt sur papier).

Sur PDA (*Personal Digital Assistant*), il existe de nombreuses applications médicales [19]. Les PDA sont en effet de plus en plus utilisés par les médecins pour leur grande mobilité, notamment à l'hôpital ou lors de consultations à domicile. En revanche, l'un des inconvénients majeurs est la petite taille de leur écran. Le langage VCM permettrait de réduire le volume des connaissances afin de les faire tenir sur ces écrans, et "Monsieur VCM" a des dimensions qui conviendraient parfaitement à l'écran des PDA. De plus, les écrans tactiles permettraient de cliquer sur les icônes simplement en posant son doigt dessus. Enfin, les difficultés techniques liées à ce type d'outil (faible mémoire disponible notamment) sont en passe d'être résolues.

8.2.4 Utilisation de VCM lors de la formation

Le langage VCM pourrait aussi être utilisé lors de la formation médicale, initiale ou continue. En effet, il est bien connu que notre mémoire retient beaucoup plus facilement les images que le texte. Le langage graphique UVAL-MED avait donné de bons résultats lors de son évaluation sur un groupe d'étudiants, cependant, contrairement à UVAL-MED, VCM n'a pas été conçu dans cette optique à l'origine.

Par ailleurs, la formation des professionnels de santé au langage VCM est une question qui reste ouverte. Cette formation peut être réalisée sous forme d'auto-formation, soit avec un didacticiel, soit directement avec les logiciels médicaux incluant VCM et disposant de fonctions d'aide, permettant par exemple de demander le sens d'une icône inconnue. La formation peut aussi être assurée au cours de journées de formation continue, voire lors de la formation initiale. Dans ce cas, une ou deux journées paraissent suffisantes pour apprendre les bases du langage (voir section 7.4.2.1).

8.2.5 Utilisation de VCM auprès d'autres professionnels de santé

Lors de notre évaluation, nous avons pris un groupe de médecins généralistes. Cependant, le langage VCM est destiné à l'ensemble des professionnels de santé. **Les spécialistes** pourraient utiliser des versions spécialisées de VCM (voir section 8.2.1).

Si le médecin n'a que peu de temps à consacrer pour rechercher des connaissances médicales, **le pharmacien** en a encore moins. Le pharmacien n'a pas accès au dossier patient, et ne dispose donc que de connaissances partielles et générales sur le patient ; par exemple il saura que untel a des problèmes cardiaques parce qu'il lui a délivré des médicaments cardiaques la semaine dernière. Dans ce cas, l'utilisation d'un langage graphique peu précis comme VCM est pertinente pour le pharmacien, qui pourra ensuite demander des précisions au patient si nécessaire. Nous pouvons même envisager d'afficher automatiquement les "Messieurs VCM" permettant la vérification des contre-indications et l'éducation du patient, lorsque les boîtes des médicaments sont passés sous la douchette.

Pour **l'infirmière**, le langage VCM permettrait de résumer de manière concise et agréable les propriétés les plus importantes d'un médicament. Ces recommandations pourraient alors être imprimées sur des aide-mémoires, ou sur des étiquettes placées sur les boîtes des médicaments, voire être directement inclus sur l'emballage du médicament. Pour l'infirmière, le langage VCM pourrait être étendu avec de nouvelles icônes indiquant le mode d'administration du médicament.

De nouvelles évaluations seront nécessaires auprès des différents types de professionnels, afin de mesurer non seulement la vitesse de lecture et le nombre d'erreurs, mais aussi la facilité d'apprentissage de VCM. Ces évaluations pourront suivre un schéma similaire à celui que nous avons utilisé sur les médecins généralistes, et le jeu de donnée que nous avons utilisé est suffisamment général pour être réutilisé. Une étape de validation en situation clinique apparaît importante.

8.2.6 Un langage graphique médical pour les patients

Le patient est aussi un acteur intéressé par les connaissances sur le médicament. Cependant, le langage VCM a été conçu en prenant en compte les connaissances dont disposent les professionnels, et ne peut donc pas être proposé au patient tel quel. C'est pourquoi nous pensons qu'il faudrait mettre au point **un second langage graphique, destiné au patient** et conçu en partant des connaissances médicales et de la "vision" de la médecine dont dispose le patient. Ce langage pourrait se rapprocher des pictogrammes de l'USP, tout en restant plus général et en disposant d'une véritable grammaire graphique. Les objectifs ne sont pas les mêmes que ceux du langage VCM pour les professionnels : pour le patient, il s'agit de donner des recommandations sur la prise du médicament, ou bien d'expliquer le mécanisme du processus pathologique et du traitement. Il est connu que, lorsque le patient est mieux informé, l'observance au traitement est meilleure [39].

L'utilisation d'approches graphiques pour transmettre des connaissances au patient présente de nombreux avantages. Un langage graphique rend les connaissances médicamenteuses accessibles aux patients qui ne savent pas lire ou ne comprennent pas la langue (jeunes enfants, touristes étrangers ou illettrés). De plus, l'aspect "attractif" des langages graphiques pourraient amener plus de patients à s'intéresser aux connaissances sur les médicaments qu'ils prennent, au contraire des notices austères que l'on trouve actuellement dans les boîtes des médicaments. En plus des langages graphiques, d'autres approches graphiques sont possibles, comme l'utilisation de jeux vidéo à des fins d'éducation du patient, par exemple pour les enfants asthmatiques ou diabétiques [92] (voir aussi une liste impressionnante de jeux : <http://www.socialimpactgames.com>, catégorie "health & wellness games").

8.3 Perspectives en sciences cognitives

8.3.1 Une méthode rigoureuse de conception des langages graphiques

Dans les chapitres 4 et 5, nous avons proposé une méthode pour la conception d'un langage graphique. Cette méthode nous a permis dans un premier temps de définir le modèle des connaissances sur lequel le langage s'appuie, à partir des connaissances d'expert, des modèles existants et d'analyses de TAL. Ensuite, nous avons construit la grammaire graphique de VCM en nous appuyant sur l'analogie et les conventions existantes, la combinatoire, la représentation graphique des relations *est-un* et la combinaison judicieuse des différents signes.

Cette méthode a été mise au point pour la conception de VCM et nous pensons qu'elle pourrait être adaptée et appliquée à d'autres domaines, par exemple en droit à la conception d'un langage graphique pour les textes de lois. Une méthode totalement rigoureuse pour la conception des langages graphiques apparaît irréalisable, en particulier à cause des aspects artistiques ou culturels qui ne peuvent être négligés. Cependant, notre méthode pourrait servir de base pour **une méthode "la plus rigoureuse possible"**.

Nous avons également proposé une méthode pour évaluer VCM au chapitre 7. Cette méthode d'évaluation pourrait aussi être appliquée dans d'autres domaines, ou à d'autres langages graphiques : en effet, les approches graphiques sont bien souvent insuffisamment évaluées, par exemple en se limitant à des questionnaires ou des interviews d'utilisateurs. En particulier, les "RCP chimériques" et la méthode pour les générer sont réutilisables dans d'autres évaluations.

8.3.2 Validation des hypothèses

Les trois hypothèses que nous avons formulées à la section 5.2 (portant sur la représentation explicite des connaissances implicites, l'utilisation de l'analogie dans le lexique, et le parallélisme entre la décomposition visuelle et la grammaire) n'ont pas été vérifiées de manière expérimentale. Seule la première a fait l'objet d'une vérification limitée à une application donnée, lors de l'évaluation de "Monsieur VCM". La validation de ces hypothèses pourrait faire l'objet d'une étude de sciences

cognitives ; cependant ces hypothèses sont de portée très générale et leur validation peut être difficile, à moins de se restreindre à une application donnée.

8.3.3 Un langage indépendant de la langue et de la culture

Nous avons vu que les langages graphiques peuvent être “universels” et donc indépendants de la langue et de la culture, grâce à l’analogie. Certains concepts abstraits (interrogation ou négation par exemple) ne peuvent être représentés de manière analogique, cependant il existe des conventions très largement connues pour les représenter (“?” pour l’interrogation ou une croix pour la négation). Si ces conventions ne sont pas “universelles”, elles sont partagées par plusieurs cultures et tendent à se répandre.

La réalisation d’un langage indépendant de la langue et de la culture française ne faisait pas partie de nos objectifs initiaux (voir 1.2), essentiellement pour des raisons pratiques : moi-même et mes encadrants, nous sommes tous de culture française et européenne. Il est donc difficile de prétendre concevoir un langage culturellement indépendant dans ces conditions ! De plus, nous avons conçu notre modèle des connaissances à partir d’une analyse d’un corpus de RCP en français, et sur les connaissances d’experts français. La langue française a donc pu avoir une influence sur le modèle.

Le langage médical varie relativement peu dans l’ensemble des langues occidentales (sans doute à cause de l’influence du latin et du grec), mais peut varier en dehors de cette limite. Par exemple, en français une classe thérapeutique est généralement nommée par la pathologie qu’elle traite, plutôt que par son activité. Il en est sans doute de même en anglais, mais peut-être pas en chinois. L’anatomie et la forme des organes sont quasi- “universels”, même si certaines cultures peuvent considérer la représentation de tout ou partie du corps humain comme “tabou”. En revanche, les pictogrammes représentant les modes de vie sont très dépendants de la culture. Par exemple le pictogramme de l’alimentation représente une assiette pleine avec des couverts, or en Asie les couverts sont remplacés par des baguettes, et en Afrique il faudrait représenter une main ! Cependant, ces pictogrammes sont peu nombreux dans VCM. **L’ “universalité” de la version actuelle de VCM n’est donc pas totale**, et cette “universalité” pourrait faire l’objet d’une évaluation dans un cadre international.

À partir de ce constat, deux solutions sont envisageables : soit mettre au point plusieurs versions de VCM adaptées aux différentes cultures, soit améliorer VCM pour arriver à un langage le plus “universel” possible. La seconde solution, si elle est plus difficile à mettre en oeuvre, nous apparaît plus intéressante : un langage graphique indépendant de toute langue et culture serait très utile pour faciliter l’accès aux connaissances médicales dans l’ensemble du monde. En effet, les textes de référence ne sont pas toujours disponibles dans la langue maternelle du médecin ; celui-ci est alors obligé de se référer à un texte dans une langue qu’il maîtrise moins bien. Si ces textes étaient enrichis avec un langage graphique, ils seraient plus facilement accessibles pour des personnes maîtrisant imparfaitement la langue.

L’amélioration de l’ “universalité” de VCM devra faire l’objet de coopérations avec des laboratoires ou des chercheurs issus de cultures différentes. Elle pourra reposer sur l’utilisation accrue de principes cognitifs universels comme l’analogie, et s’inspirer des travaux qui ont été réalisés pour la conception des langages graphiques “universels” (voir section 2.2.1.3).

8.4 Perspectives en informatique et en représentation des connaissances

8.4.1 Traduction automatique de textes en VCM

La traduction des documents textuels existants en VCM reste un point délicat. Afin de faciliter la traduction, notre prototype utilise un langage XML pour structurer les connaissances médicales ; il suffit de baliser le texte selon ce langage pour pouvoir ensuite le traduire automatiquement en VCM ou en “Monsieur VCM”. Cependant le balisage des textes reste problématique. Pour l’instant

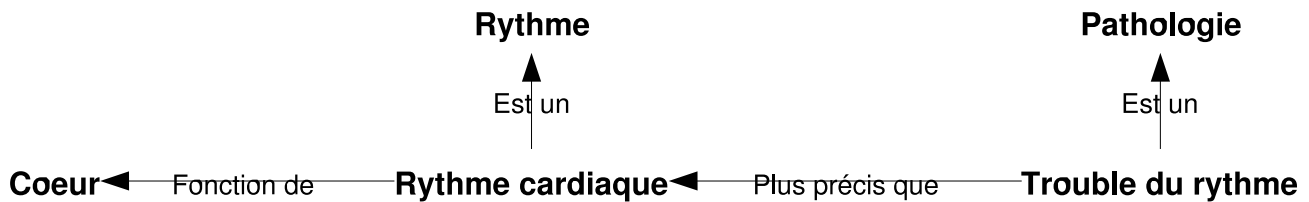


Fig. 8.1 – Exemple de représentation d’une partie d’une terminologie sous forme de réseau. L’exemple est ici appliquée au concept “trouble du rythme”; à comparer avec l’icône VCM correspondante.

les documents en VCM dont nous disposons ont soit été balisés à la main, soit proviennent de connaissances déjà structurées extraites de bases comme Thériaque. En ce qui concerne les Résumés des Caractéristiques Produits ou les Guides de Bonnes Pratiques, des versions structurées peuvent être disponibles. En revanche, si nous souhaitons représenter en VCM par exemple un compte rendu d’hospitalisation, cela n’est pas forcément le cas.

La traduction automatique de texte libre en VCM passe par le balisage automatique du texte libre. Ce balisage automatique pourrait être réalisé en utilisant des méthodes de Traitement Automatique du Langage (TAL). Le langage XML utilisé, tout comme VCM lui-même, est peu précis, ce qui facilite la tâche. Des méthodes de TAL spécifiques pour l’encodage “grossier” de connaissances pourraient être mises au point.

Des méthodes semi-automatiques peuvent aussi être envisagées. Par exemple, la reconnaissance des pathologies et des médicaments dans un texte est aisée et déjà bien maîtrisée, et donc facile à automatiser. Quelle que soit la méthode choisie, une vérification humaine sera nécessaire afin de garantir l’absence d’erreur dans le document VCM; des outils spécifiques destinés à faciliter cette vérification sont aussi envisageables, à la manière de DOPAMINE pour la Galen Drug Ontology (voir section 2.3.3.4).

8.4.2 Génération de traducteurs à partir de la grammaire graphique

Nous avons proposé un formalisme pour représenter la grammaire d’un langage graphique (voir section 5.2.6). Actuellement, la traduction des connaissances médicales structurées en icônes VCM est réalisée de manière automatique par un logiciel écrit spécifiquement pour la grammaire de la version actuelle de VCM. Nous pourrions aller plus loin, en générant de manière automatique ce programme à partir de la description formelle de la grammaire graphique. Cela faciliterait la mise au point des nouvelles versions de VCM, car il suffirait de formaliser une nouvelle grammaire pour pouvoir l’utiliser dans notre prototype. De plus, le générateur de traducteur serait indépendant de VCM, et il pourrait donc servir à mettre au point d’autres langages graphiques, dans n’importe quel domaine.

Comme le formalisme que nous avons proposé est un formalisme graphique s’appuyant sur UML, cela pourrait aboutir à **un environnement graphique de conception de langages graphiques**. Dans cette environnement, l’utilisateur pourrait dessiner le modèle UML des connaissances à représenter, puis le modèle UML des éléments graphiques du langage, et enfin relier les classes des deux modèles par des relations “est représenté par”. Ensuite, un programme serait généré automatiquement pour traduire des connaissances structurées (par exemple au format XML) vers le langage graphique défini par l’utilisateur.

8.4.3 Représentation graphique de terminologies ou d’ontologies

Les terminologies permettent de classer des concepts, et les ontologie comprennent la définition de l’ensemble des concepts d’un domaine. Actuellement, les terminologies ou les ontologies ne sont que rarement présentées au médecin ou à l’utilisateur final d’un système : elles sont jugées trop complexes

et / ou trop abstraites pour cela.

Lorsqu'elles sont représentées graphiquement, c'est **sous forme de réseaux** montrant les concepts et les relations entre eux (voir exemple figure 8.1). Cependant, cette représentation pose un problème : **il n'est pas possible de représenter un concept seul**. Par exemple, pour indiquer la relation entre "rythme cardiaque" et "trouble du rythme", nous sommes obligés de représenter les deux concepts. Si certaines relations sont héritées *via* un concept parent, nous sommes obligés de représenter plus de deux niveaux de concept, et ainsi de suite. Cependant, en particulier dans le cas d'ontologies de grande taille comme UMLS (*Unified Medical Language System*), il n'est pas possible de représenter la totalité des concepts.

Il semble possible d'utiliser **un langage graphique pour représenter les concepts d'une terminologie ou d'une ontologie** par des icônes. Ces icônes peuvent indiquer tout ou partie des attributs du concept qu'elle représente, ainsi que des relations entre le concept représenté et d'autres concepts; nous avons vu comment représenter les relations *est-un* à la section 5.2.3. Si l'on prend l'icône VCM pour "trouble du rythme", celle-ci exprime sensiblement les mêmes informations que le réseau de la figure 8.1; cependant les différents concepts (coeur, pathologie, rythme,...) sont combiné dans une seule icône, ce qui réduit considérablement le volume occupé (dans notre exemple, une icône au lieu de cinq groupes nominaux et quatre relations).

L'utilisation de langages graphiques pour représenter des concepts et les relations entre eux apparaît donc comme une alternative à la visualisation en réseau. Cette méthode est particulièrement adaptée lorsque l'on souhaite représenter un grand nombre de concepts. Elle pourrait rendre possible en pratique la présentation d'une terminologie ou d'une ontologie à l'utilisateur final, ce qui ouvre de nouvelles perspectives qui restent à développer.

8.4.4 Visualisation et recherche d'information ou de connaissance ?

Dans le domaine de la recherche d'information [153] ou la visualisation d'information (voir section 2.3, *information retrieval* et *information visualisation* en anglais), la littérature est abondante. En revanche, elle l'est beaucoup moins lorsque l'on s'intéresse à la recherche de connaissance ou la visualisation de connaissance (*knowledge retrieval* et *knowledge visualisation* en anglais).

La visualisation de connaissance est un domaine de recherche extrêmement récent. R. Burkhard [21] le définit comme l'utilisation d'approche graphique pour faciliter le transfert de connaissance d'une personne à une autre. Cependant cette définition paraît trop restrictive : la visualisation de connaissances peut avoir plusieurs objectifs différents :

- **Le transfert de connaissance et l'apprentissage** (situation dans laquelle R. Burkhard se place, que l'on retrouve pour UVAL-MED ou les pictogrammes de l'USP).
- **La mise en application immédiate des connaissances visualisées** (par exemple la prise en compte d'une contre-indication par le médecin pour un patient donné, en utilisant VCM, ou la recherche d'incompatibilité entre solutions injectables avec Stabilis 2). D'autres auteurs ont proposé des outils visualisant des connaissances pour faciliter leur mise en application, par exemple pour visualiser sous forme d'arbre les différentes procédures d'urgence à exécuter en cas de problèmes sur les voies ferrées [115].
- **L'écriture, l'enrichissement ou la vérification des connaissances visualisées** (c'est par exemple le cas de DOPAMINE [161], voir section 2.3.3.4).

Si nous avons vu que la connaissance est une catégorie d'information (section 3.1), je pense néanmoins (tout comme R. Burkhard) que la visualisation de connaissance nécessite des méthodes et des techniques différentes de la visualisation d'information, pour les raisons suivantes :

- **La connaissance est généralement de nature différente de l'information**. Si l'information peut être quantitative, la connaissance est très souvent qualitative. La connaissance peut aussi se présenter sous forme de règles si-alors ou de relations cause-conséquence ou *est-un*.
- **La connaissance n'est pas utilisée de la même manière que l'information**. La recherche ou la visualisation d'information ont pour objectif soit l'utilisation directe de l'information trouvée (par exemple lors d'une recherche bibliographique), soit l'extraction de connaissance

(*data-mining* par exemple). Au contraire, nous avons vu que la recherche ou la visualisation de connaissances ont pour objectif l'apprentissage, la mise en application des connaissances, ou l'écriture et la vérification des connaissances.

Les outils de recherche ou de visualisation de connaissance doivent donc, selon leurs utilisations, être optimisés pour faciliter l'apprentissage (par exemple par l'emploi de procédés pédagogiques et mnémotechniques), pour faciliter une application donnée des connaissances (par exemple en représentant de manière explicite les connaissances implicites utiles à cette application ; c'est ce que nous avons fait avec "Monsieur VCM", qui est optimisé pour la vérification des contre-indications), ou pour faciliter l'écriture et la vérification des connaissances (par exemple en donnant une vue d'ensemble des connaissances comme le fait DOPAMINE).

8.5 Conclusion

Notre objectif initial était de mettre au point des approches graphiques adaptées à la visualisation des connaissances sur le médicament, permettant aux professionnels de santé d'accéder plus rapidement aux connaissances sur le médicament. Pour répondre à cela, nous nous sommes placés dans un cadre multi-disciplinaire : médecine et pharmacie, sciences cognitives et informatique, et nous avons proposé dans cette thèse plusieurs éléments méthodologiques :

- une méthode de TAL pour extraire, à partir d'un corpus de texte, la fréquence des attributs définissant les principaux concepts, et les principaux motifs de phrases-types (section 4.2.3.2 et 4.2.3.3),
- une méthode pour mettre au point un modèle simplifié des connaissances, à partir des connaissances d'expert, des modèles existants et d'un corpus de texte (chapitre 4),
- un ensemble de règles pour mettre au point une grammaire de langage graphique, en prenant en compte les capacités de la vision humaine (chapitre 5),
- un formalisme basé sur UML pour représenter une grammaire graphique (section 5.2.6),
- une méthode pour indexer graphiquement un document médical en fonction des localisations anatomico-fonctionnelles et des étiologies, "Monsieur VCM" (chapitre 6),
- une méthode pour générer des "documents chimériques" à partir d'un corpus de documents (section 7.2.6),
- une méthode pour comparer, dans des conditions contrôlées, une approche graphique à un texte utilisé dans le but de présenter des connaissances (chapitre 7).

En s'appuyant sur ces éléments méthodologiques, nous avons pu réaliser :

- **le langage graphique VCM** (Visualisation des Connaissances Médicales), disponible dans une version stable 1.0,
- "**Monsieur VCM**", un outil graphique permettant d'indexer graphiquement un document en fonction des localisations anatomico-fonctionnelles et des étiologies,
- un matériel pédagogique (manuel et didacticiel) pour l'apprentissage de VCM,
- un prototype de base de connaissance exploitant VCM et "Monsieur VCM", et utilisant les connaissances de la base Thériaque,
- l'évaluation de VCM et de "Monsieur VCM" sur un petit groupe de médecins généralistes, qui a permis de montrer que, dans des conditions contrôlées, **les médecins ont en moyenne lu VCM deux fois plus vite que le texte, en faisant deux fois moins d'erreurs.**

Notre objectif initial, accélérer l'accès aux connaissances médicales pour les professionnels de santé, semble donc atteint. De plus, nos résultats montrent une diminution du nombre d'erreurs qui n'était pas attendue. VCM pourrait donc faciliter l'accès au connaissance médicale en clinique, mais aussi **sécuriser** cet accès.

Les approches graphiques nous semblent donc très prometteuses pour présenter les connaissances médicales aux professionnels de santé, dans le but d'améliorer la vitesse de lecture, la compréhensibilité et l'attractivité de ces connaissances. Ces approches devraient être étendues à l'ensemble du domaine médical, et des approches spécifiques pourraient être développées pour le patient. En ce qui concerne le langage VCM, les perspectives incluent l'extension du langage à l'ensemble de la médecine

et l'amélioration de son "universalité" vis-à-vis des différentes cultures et son utilisation dans d'autres applications médicales comme le dossier patient.

Bibliographie

- [1] *L'aventure des écritures – dossier de la BnF*. <http://classes.bnf.fr/dossiecr/index.htm>.
- [2] 'Recommandation de 1999 pour le traitement de l'hypertension artérielle au Canada', *Journal de l'Association Médicale Canadienne* **161** (1999), no. 12suppl. <http://www.chs.md/pdf/hyper-f.pdf>.
- [3] 'Les Québécois et les pharmaciens : sondage d'opinion', *Tech. Report* (Ordre des pharmaciens du Québec, Jul. 2002). http://www.opq.org/fr/services_membres/.
- [4] AFSSAPS, *Médicaments et conduite automobile : de nouveaux pictogrammes plus informatifs* (Communiqué de presse, 22 Sep. 2005). <http://agmed.sante.gouv.fr/htm/10/filcoprs/050905.htm>.
- [5] C. Ahlberg, C. Williamson, and B. Shneiderman, 'Dynamic queries for information exploration : An implementation and evaluation', in *Proc. ACM Conf. Human Factors in Computer Systems, CHI* (1992), 619–626. <http://citeseer.nj.nec.com/alhlberg92dynamic.html>.
- [6] K. Andrews, *Information visualisation : tutorial notes* (5 Jul. 2002). <http://www2.iicm.edu/ivis/ivis.pdf>.
- [7] M. Ash, *Gestalt Psychology In German Culture 1890 - 1967* (Cambridge University Press, Cambridge, 1995).
- [8] P. Barr, J. Noble, and R. Biddle, 'Icons r icons', in *Proceedings of the Fourth Australian user interface conference on User interfaces 2003* (Australian Computer Society, Inc., 2003), 25–32. <http://crpit.com/Vol18.html>.
- [9] B. Bederson and J. Hollan, *Pad++ : A zoomable graphical interface*, in *Proceedings of ACM Human Factors in Computing Systems Conference Companion (CHI 95)* (1995). <http://citeseer.nj.nec.com/35138.html>.
- [10] C. Benson, A. Elman, S. Nickell, and C. Robertson, *GNOME Human Interface Guidelines (1.0)* (20 Aug. 2002). <http://developer.gnome.org/projects/gup/hig/1.0/>.
- [11] D. Berry, I. Michas, and T. Gillie, 'What do patients want to know about their medicines and what do doctors want to tell them', *Psychology and health* **12** (1997), 467–480.
- [12] J. Bertin, *Sémiologie graphique : Les diagrammes - Les réseaux - Les cartes* (Editions de l'Ecole des Hautes Etudes en Sciences, Paris-La Haye, 1999).
- [13] ———, *Sémiologie graphique* (2001). <http://www.sciences-po.fr/cartographie/semio/>.
- [14] O. Bodenreider and A. McCray, 'Exploring semantic groups through visual approaches', *Journal of biomedical informatics* **36** (2003), 414–432.
- [15] C. Bouthier, *Two free tools for tree-data visualization*, in *Journée Visualisation de l'Information* (URI (INIST-CNRS) - CORTEX (INRIA-LORIA), 19 Feb. 2003). <http://eicstes.inist.fr/public/Bouthier.pdf>.
- [16] K. Boyack, B. Wylie, G. Davidson, and D. Johnson, *Analysis of patent databases using vxinsight*, in *Proceedings of the 9th conference on information and knowledge management (CIKM 2000)* (McLean, VA, 10 Nov. 2000). <http://citeseer.nj.nec.com/581895.html>.
- [17] E. Brown and E. Clark, 'Evaluation of MedDRA in representing medical product data sheet information', *Pharmaceutical medicine* (1996), no. 10, 111–118.

- [18] C. Bruley, *Analyse des représentation graphique de l'information – extension aux représentations tridimensionnelles* (Ph.D. thesis, Université Joseph Fourier, 3 Jun. 1999). <ftp://ftp.inrialpes.fr/pub/romans/theses/bruley.ps.gz>.
- [19] P. Brunet and P. Le Beux, *L'étudiant en Médecine et l'Assistant Numérique Personnel dans le cadre de l'Université Médicale Virtuelle Francophone*, in *Journées Francophones d'Informatique Médicale (JFIM2005)* (Lille, 13 May 2005).
- [20] A. Burgun, O. Bodenreider, and F. Mougin, 'Classifying diseases with respect to anatomy : a study in SNOMED CT.', *AMIA Annu Symp Proc* (2005), 91–5 (eng).
- [21] R. Burkhard, *Knowledge visualization - The use of complementary visual representations for the transfer of knowledge. A model, a framework, and four new approaches* (Ph.D. thesis, Swiss Federal Institute of Technology (ETH Zurich), Zürich, 2005). <http://e-collection.ethbib.ethz.ch/show?type=diss&nr=15918>.
- [22] E. Campbell, D. Sittig, J. Ash, K. Guappone, and R. Dykstra, 'Types of unintended consequences related to computerized provider order entry', *Journal of the American medical informatics association* **13** (2006), no. 5, 547–556.
- [23] S. K. Card and J. Mackinlay, 'The structure of the information visualization design space', in *Proceedings of Information Visualization '97, IEEE* (1996), 92–99. <http://citeseer.nj.nec.com/card96structure.html>.
- [24] M. Cartier, *Les icônes à l'écran - Une sténographie de l'interactivité, Laboratoire de télématique, Université du Québec à Montréal (UQAM)* (Dec. 2004). http://www.michelcartier.com/McArticleB.php3?id_article=374.
- [25] S. Chang, *Visual Languages*, ch. Iconic Visual Languages, pp. 1–7 (Plenum Press, New York, 1986). <http://www.cs.pitt.edu/~%7Echang/365/sk1.html\#examples>.
- [26] E. Chazard, P. Puech, M. Gregoire, and R. Beuscart, 'Using Treemaps to Represent Medical Data', *Studies in Health Technology and Informatics* **124** (2006), 522–527.
- [27] H. Chernoff, 'Using faces to represent points in k-dimensional space graphically', *Journal of the American Statistical Association* **68** (1973), no. 342, 361–368.
- [28] Chi E.H., 'A taxonomy of visualization techniques using the data state reference model', in *Proceedings of the Symposium on Information Visualization (InfoVis '00)* (IEEE Press, Salt Lake City, Utah, 2000), 69–76.
- [29] L. Chittaro, 'Information visualization and its application to medicine', *Artificial Intelligence in Medicine* **22** (2001), no. 2, 81–88. <http://citeseer.nj.nec.com/chittaro01information.html>.
- [30] L. Chittaro, C. Combi, and G. Trapasso, *Visual data mining of clinical databases : An application to the hemodialytic treatment based on 3d interactive bar charts* (2000). <http://citeseer.nj.nec.com/chittaro02visual.html>.
- [31] M. Chuah and S. Eick, 'Glyphs for software visualization', in *International Workshop on Program Comprehension* (Dearborn, MI, USA, May 1997), 183–191. <http://citeseer.nj.nec.com/chuah97glyphs.html>.
- [32] W. Cole and J. Stewart, 'Human performance evaluation of a metaphor graphic display for respiratory data', *Methods of information in medicine* **33** (1994), no. 4, 390–396.
- [33] S. Crawford and S. Dombrowski, 'Extemporaneous compounding activities and the associated informational needs of pharmacists', *American journal of hospital pharmacy* **48** (1991), no. 6, 1205–1210.
- [34] A. Dadzie and A. Burger, 'Providing visualisation support for the analysis of anatomy ontology data', *BMC Bioinformatics* **6** (2005), no. 74. <http://www.biomedcentral.com/1471-2105/6/74>.

- [35] L. Dazzi, C. Fassino, R. Saracco, S. Quaglini, and M. Stefanelli, *A patient workflow management system built on guidelines*, in *Proceedings of the AMIA annual Fall Symposium* (1997). <http://www.amia.org/pubs/symposia/D004083.PDF>.
- [36] B. De Carolis, F. De Rosis, and S. Errore, ‘A user-adapted iconic language for the medical domain’, *International Journal of Human-Computer Studies* **43** (1995), no. 4, 561–577.
- [37] S. Dewally, *Transcription des précautions d’emploi en pictogrammes, évaluation par le public* (thèse de pharmacie, UFR de pharmacie, Université Joseph Fourier, Grenoble, France, 1998).
- [38] D. Dickinson, T. Raynor, J. Kennedy, S. Bonaccorso, and J. Sturchio, ‘What information do patients need about medicines?’, *British Medical Journal* **327** (2003), 861–864. <http://bmj.bmjournals.com/cgi/reprint/327/7419/861-a.pdf>.
- [39] A. Dolovich, ‘Répondre aux besoins des patients en matière d’information sur les médicaments’, *Journal Canadien de pharmacie hospitalière* **56** (2003), no. 3, 137–139. http://www.cshp.ca/our_journal/Juneissue2003JRL/june.ED.fre.pdf.
- [40] C. Dos Santos, *Visualisation métaphorique tridimensionnelle de l’information* (Ph.D. thesis, Ecole nationale supérieure des télécommunications (PARIS-ENST), 2002).
- [41] P. Dourgnon, N. Grandfils, M. Sourty-Le Guellec, and M. Zimina, ‘Apport de l’informatique dans la pratique médicale – étude FORMMEL’, (2000). <http://www.credes.fr/Formmel/rapport.pdf>.
- [42] H. Dreyfuss, *Symbol sourcebook* (1984).
- [43] C. Duclos, *Représentation de l’information pharmaco-thérapeutique des résumé des caractéristiques produit des médicaments : apport des méthodes de traitement du langage naturel, développement, validation et utilisation de modèles* (Ph.D. thesis, Université de Paris 5 – René Descartes – Faculté de médecine Cochin Port Royal, 25 Apr. 2003).
- [44] C. Duclos, G. Cartolano, M. Ghez, and A. Venot, ‘Structured representation of the pharmacodynamics section of the Summary of Product Characteristics for antibiotics : application for automated extraction and visualization of their antimicrobial activity spectra’, *Journal of the American Medical Informatics Association* **11** (2004), no. 4, 285–293.
- [45] C. Duclos and A. Venot, ‘Structured representation of drug indications : lexical and semantic analysis and object-oriented modeling’, *Methods of information in medicine* **39** (2000), 83–87.
- [46] C. Duclos-Cartolano and A. Venot, ‘Building and evaluation of a structured representation of pharmacokinetics information presented in SPCs : from existing conceptual views of pharmacokinetics associated with natural language processing to object-oriented design.’, *Journal of the American Medical Informatics Association* **10** (2003), no. 3, 271–80 (eng).
- [47] V. Ebrahiminia, C. Duclos, M. Toussi, C. Riou, R. Cohen, and A. Venot, ‘Representing the Patient’s Therapeutic History in Medical Records and in Guideline Recommendations for Chronic Diseases Using a Unique Model.’, *Studies in Health Technology and Informatics* **116** (2005), 101–6 (eng).
- [48] V. Ebrahiminia, C. Riou, B. Seroussi, J. Bouaud, S. Dubois, H. Falcoff, and A. Venot, ‘Design of a Decision Support System for Chronic Diseases Coupling Generic Therapeutic Algorithms with Guideline-Based Specific Rules’, *Studies in Health Technology and Informatics* **124** (2006), 483–488.
- [49] S. Eick, ‘Graphically displaying text’, *Journal of Computational and Graphical Statistics* **3** (1994), no. 2, 127–142. <http://citeseer.nj.nec.com/eick94graphically.html>.
- [50] EISTI, *Qu’est-ce que la sémiotique ?* (1999). <http://www.eisti.fr/res/doc/semiotique/>.
- [51] L. Elting and G. Bodey, ‘Is a picture worth a thousand medical words? A randomized trial of reporting formats for medical research data’, *Methods of information in medicine* **30** (1991), 145–150.

- [52] L. Elting, C. Martin, S. Cantor, and E. Rubenstein, ‘Influence of data display formats on physician investigators’ decisions to stop clinical trials : prospective trial with repeated measures.’, *British Medical Journal* **318** (1999), no. 7197, 1527–1531 (eng). <file:///home/jiba/bmj.com/cgi/content/full/319/7216/1070>.
- [53] J. Ely, J. Osheroff, M. Ebell, G. Bergus, B. Levy, M. Chambliss, and E. Evans, ‘Analysis of questions asked by family doctors regarding patient care’, *British Medical Journal* **319** (1999), 358–361. <http://bmj.bmjournals.com/cgi/content/full/319/7206/358>.
- [54] J. Ely, J. Osheroff, M. Ebell, M. Chambliss, D. Vinson, J. Stevermer, and E. Pifer, ‘Obstacles to answering doctors’ questions about patient care with evidence : qualitative study’, *British Medical Journal* **324** (2002), no. 7339, 710. <http://bmj.bmjournals.com/cgi/content/full/324/7339/710>.
- [55] R. Erbacher, ‘Glyph-based generic network visualization’, in *Proceedings of the SPIE ’2002 conference on Visualization and Data Analysis* (2002), 228–237.
- [56] K. P. Fishkin and M. C. Stone, ‘Enhanced dynamic queries via movable filters’, in *CHI* (1995), 415–420. <http://citeseer.nj.nec.com/fishkin95enhanced.html>.
- [57] P. Frame, ‘Automated Health Maintenance Reminders : Tools Do Not Make a System’, *Journal of the American Board of Family Practice* **16** (2003), 350–351. <http://www.jabfm.org/cgi/content/full/16/4/350>.
- [58] FULMEDICO and Y. Adenis-Lamarre, *Enquête sur les bases de données médicamenteuses électroniques* (2000). http://www.stethonet.org/enquete_bdm/index.htm.
- [59] G. Furnas and J. Zacks, *Multitrees : Enriching and reusing hierarchical structure*, in *Human Factors in Computing Systems : proceedings of the CHI ’94 Conference* (New York, 1994). <http://citeseer.nj.nec.com/furnas94multitrees.html>.
- [60] G. Furnas, ‘Generalized Fisheye views’, in *Proceedings of the Human Factors in Computing Systems CHI ’86 conference* (1986), 16–23. <http://www.si.umich.edu/~furnas/Papers/FisheyeCHI86.pdf>.
- [61] M. Graham, J. Kennedy, and D. Benyon, ‘Towards a methodology for developing visualizations’, *International journal of Human-computer studies* **53** (2000), 789–807. <http://www.soc.napier.ac.uk/publication/op/getpublication/publicationid/230055>.
- [62] M. Graham, *Visualising multiple overlapping classification hierarchies* (Ph.D. thesis, Napier, Dec. 2001). <http://citeseer.nj.nec.com/article/graham01visualising.html>.
- [63] M. Green, *Toward a perceptual science of multidimensional data visualization : Bertin and beyond* (1999). <http://ergogero.com/dataviz/dviz0.html>.
- [64] A. Guarnero, M. Marzuoli, G. Molino, P. Terenziani, M. Torchio, and K. Vanni, ‘Contextual and temporal clinical guidelines’, *Proceedings of the AMIA Symposium* (1998), 683–687. <http://www.amia.org/pubs/symposia/D005195.PDF>.
- [65] K. Hämeen-Anttila, K. Kemppainen, H. Enlund, J. Patricia, and A. Marja, ‘Do pictograms improve children’s understanding of medicine leaflet information?’, *Patient Education and Counseling* **55** (2004), 371–378.
- [66] M. Hascoët and M. Beaudouin-Lafon, ‘Visualisation interactive d’information’, *Information, interaction, intelligence (I3)* **1** (2001), no. 1, 77–108. <http://www.lirmm.fr/~mountaz/Publi/revueI3-hascoet.pdf>.
- [67] J. Haug, ‘Physicians’ preferences for information sources : a meta-analytic study’, *Journal of the Medical Library Association* **85** (1997), no. 3, 223–232. <http://www.pubmedcentral.nih.gov/pagerender.fcgi?artid=226263>.
- [68] M. Hearst, *Modern Information Retrieval*, ch. User interfaces and visualization, pp. 257–322 (Addison Wesley Longman, 1999). <http://www.sims.berkeley.edu/~hearst/irbook/>.

- [69] K. Helbing, J. Jenkins, Y. Kim, and M. Miller, *Influence of Icon Detail, Color, and Perspective on Preference, Recognition Time, and Search Time*, in *Proceedings of Interface '93* (1993). <http://www.frontiernet.net/~mkmiller/Prof/Icons/icon.htm>.
- [70] R. Hendley, N. Drew, A. Wood, and R. Beale, 'Narcissus : Visualising information', in *Proc. IEEE Symp. Information Visualization, InfoVis*, Eds. N. D. Gershon and S. Eick (IEEE Computer Soc. Press, 30-31 1995), 90–96. <http://citeseer.nj.nec.com/hendley95narcissus.html>.
- [71] I. Herman, G. Melançon, and M. S. Marshall, 'Graph visualization and navigation in information visualization : A survey', *IEEE Transactions on Visualization and Computer Graphics* **6** (2000), no. 1, 24–43. <http://citeseer.nj.nec.com/herman00graph.html>.
- [72] F. Holly, *XML and Pharmaceuticals : Regulatory and Non-Regulatory Applications*, in *Proceedings of XML 2002* (2002). http://www.idealliance.org/papers/xml02/dx_xml02/papers/03-04-04.
- [73] P. Houts, C. Doak, L. Doak, and M. Loscalzo, 'The role of pictures in improving health communication : a review of research on attention, comprehension, recall and adherence', *Patient education and counseling* **61** (2006), 173–190.
- [74] D. Hubel, *Eye, Brain, and Vision* (Jan. 1995). <http://neuro.med.harvard.edu/site/dh>.
- [75] K. Hunt, *Reflections on an Icon Development Process : Negotiating Design Issues*, in *STC Proceedings* (1996). <http://www.stc.org/proceedings/ConfProceed/1996/PDFs/P432435.PDF>.
- [76] T. Ingen-Housz, 'The Elephant's Memory : in search of a pictorial language', *Learning Technology Review* (1999). <http://www.apple.com/education/LTReview/spring99/elephant/index.html>.
- [77] B. Johnson and B. Shneiderman, 'Treemaps : a space-filling approach to the visualization of hierarchical information structures', in *Proceedings of the 2nd International IEEE Visualization Conference* (San Diego, Apr. 1991), 284–291.
- [78] P. Johnson, S. Tu, N. Booth, and et al, *Using scenarios in chronic disease management guidelines for primary care*, in *Proceedings of the AMIA annual Symposium* (2000). <http://www.amia.org/pubs/symposia/D200629.PDF>.
- [79] B. Ketan, 'Using Treemaps to Visualize Gene Ontologies', *Tech. Report* (Human Computer Interaction Lab and Institute for Systems Research, University of Maryland, College Park, MD USA, 12 Apr. 2001). <http://www.cs.umd.edu/hcil/treemap/GeneOntologyTreemap.pdf>.
- [80] E. Kleiberg, H. van de Wetering, and J. van Wijk, 'Botanical visualization of huge hierarchies', in *Proceedings of IEEE InfoViz* (San Diego, California, 2001), 87–94. <http://citeseer.nj.nec.com/496378.html>.
- [81] S. Kolstrup, *The semiotics of visual communication*. http://inv.au.dk/semiotics/modul_3/.
- [82] G. Kwok-Chu Ng, *Interactive visualisation techniques for ontology development* (Ph.D. thesis, University of Manchester, Nov. 2000).
- [83] G. Lairy, 'La satisfaction des patients lors de leur prise en charge dans les établissements de santé', *Tech. Report* (ANAES, Sep. 1996). <http://www.anaes.fr/anaes/Publications.nsf/>.
- [84] J.-B. Lamy, *Télé-surveillance médicale : conception d'un système d'analyse temporelle, validation par simulation et reconstitution par réalité virtuelle* (Mémoire de 5ème année de pharmacie, option recherche, 31 May 2002).
- [85] J.-B. Lamy, C. Duclos, V. Rialle, and A. Venot, 'Quelles approches graphiques pour faciliter la diffusion des connaissances médicales?', in *5ème congrès Internet et Pédagogie Médicale (IPM 2004)* (Grenoble, 3 Dec. 2004), 39–40.
- [86] ———, 'Classification et analyse critique des approches graphiques appliquées à la visualisation des connaissances médicales', in *Journées Francophones d'Informatique Médicale (JFIM2005)* (Lille, 13 May 2005), 1–9.

- [87] ———, ‘Which graphical approaches should be used to represent medical knowledge?’, *Studies in health technology and informatics* **116** (2005), 719–724.
- [88] ———, *Vers une méthode rigoureuse de conception des langages graphiques s’appuyant sur les sciences cognitives*, in *4ème atelier visualisation et extraction de connaissances, 6èmes journées francophones d’Extraction et de Gestion des Connaissances*, Eds. F. Poulet and B. Le Grand (Lille, France, 17 Jan. 2006).
- [89] M. Lanzenberger, *The interactive stardinates - design considerations*. <http://citeseer.nj.nec.com/579605.html>.
- [90] H.-Y. Lee, H.-L. Ong, and L.-H. Quek, ‘Exploiting visualization in knowledge discovery’, in *Knowledge Discovery and Data Mining* (1995), 198–203. <http://citeseer.nj.nec.com/lee95exploiting.html>.
- [91] N. Leemans, *VIL : A Visual Inter Lingua* (Ph.D. thesis, Worcester polytechnic institute, Apr. 2001). <http://www.cs.wpi.edu/~nemleem/finaldissertation.pdf>.
- [92] D. Lieberman, ‘Management of chronic pediatric diseases with interactive health games : Theory and research findings’, *Journal of Ambulatory Care Management* **24** (2001), 26–38.
- [93] J. Liu, C. Milstein, B. Séné, and A. Venot, ‘Object-oriented modeling and terminologies for drug contraindications’, *Methods of information in medicine* **37** (1998), 45–52.
- [94] D. Lobach, ‘Electronically distributed, computer-generated, individualized feedback enhances the use of a computerized practice guideline.’, *Proc AMIA Annu Fall Symp* (1996), 493–7 (eng).
- [95] J. Mackinlay, G. Robertson, and S. Card, ‘The perspective wall : detail and context smoothly integrated’, in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI ’91)* (1991), 173–179.
- [96] H. Maurer, R. Stubenrauch, and D. Camhy, ‘Foundations of MIRACLE : Multimedia Information Repository, A Computer-supported Language Effort’, *Journal of Universal Computer Science* **9** (2003), no. 4, 309–348. http://www.jucs.org/jucs_9_4/foundations_of_miracle_multimedia/.
- [97] J.-G. Meunier, ‘La structure générique des systèmes sémiotiques’, *Recherche sémiotique / Semiotic inquiries (RSSI)* (1988), no. 8.
- [98] ———, ‘The Categorial Structure of Iconic Languages’, *Theory & Psychology* **8** (1998), no. 6, 805–825. <http://www.psych.ucalgary.ca/thpsyc/VOLUMES.SI/1998/8.6.Meunier.html>.
- [99] J. Meyer, K. Perlin, B. Bederson, and J. Hollan, *Two document visualization techniques for zoomable interfaces* (1995). <http://citeseer.nj.nec.com/2806.html>.
- [100] J. Moline, ‘Virtual Reality for Health Care : a survey’, *Tech. Report* (National Institute of Standards and Technology, Gaithersburg, Maryland, 1997). <http://utenti.tripod.it/dualband/pdf/moline.pdf>.
- [101] K. Morik and M. Wurst, *Knowledge discovery and knowledge visualization*, in *Netzspannung Workshop 2002* (2002). <http://citeseer.nj.nec.com/565798.html>.
- [102] F. Nake, ‘Data, Information, and Knowledge’, in *Organizational Semiotics : Evolving a Science of Information Systems*, Eds. K. Liu, R. Clarke, P. Andersen, and R. Stamper (Kluwer, Montréal, Québec, Canada, Jul. 2001), 41–50.
- [103] F. Nunez, *An extended spreadsheet paradigm for data visualisation systems, and its implementation* (Master’s thesis, University of Cape Town, 2002). <http://citeseer.nj.nec.com/nunez02extended.html>.
- [104] A. Olsen, R. Korfhage, K. Sochats, M. Spring, and J. Williams, ‘Visualization of a document collection : the Vibe system’, *Information processing & management* **29** (1992), no. 1, 69–81. <http://l13.exp.sis.pitt.edu/Website/Webresumé/VIBEPaper/VIBE.htm>.
- [105] A. O’Rourke, *Patient satisfaction*, in *Wisdom-Trent NHSE virtual conference on clinical governance* (9 Sep. 2000). <http://www.shef.ac.uk/uni/projects/wrp/cgptsat.htm>.

- [106] N. Osawa, 'Visualization of inheritance relationships using glyphs', *IEICE Trans. on Information and Systems* **E85-D** (2002), no. 1, 275–282.
- [107] A. Paivio, *Mental representations - A dual coding approach* (Oxford University Press, New York, 1986). <http://tip.psychology.org/paivio.html>.
- [108] T. Panas, R. Berrigan, and J. Grundy, *A 3D metaphor for software production visualization*, in *7th International Conference on Information Visualization (IV03)* (London, England, Jul. 2003). <http://www.masda.vxu.se/~tps/publications/PBG-IV-03-ProductVis.pdf>.
- [109] G. Paul, *La psychologie de la forme* (Flammarion, Paris, 1979).
- [110] M. Peleg, A. Boxwala, O. Ogunyemi, Q. Zeng, S. Tu, R. Lacson, E. Bernstam, N. Ash, P. Mork, L. Ohno-Machado, E. Shortliffe, and R. Greenes, *GLIF3 : the evolution of a guideline representation format*, in *Proceedings of the AMIA Fall Symposium* (2000). <http://citeseer.nj.nec.com/peleg00glif.html>.
- [111] M. Peleg, A. Boxwala, S. Tu, O. Ogunyemi, Q. Zeng, and D. Wang, 'Guideline Interchange Format technical specification', *Tech. Report* (30 Apr. 2001). <http://smi-web.stanford.edu/projects/intermed-web/guidelines/>.
- [112] D. Peraya, 'Structures et fonctionnement sémiotiques des icônes de logiciels et d'environnements informatiques standardisés (ILEIS)', *Recherches en communication* (1998), no. 10. <http://www.comu.ucl.ac.be/reco/grems/jpweb/peraya/PerayaReC10.pdf>.
- [113] D. Peraya and D. Strasser, *Iconomètre : un outil de formation et de recherche pour la mesure du potentiel évocateur des représentations visuelles*, in *5ème Colloque Européen sur l'Autofor-mation* (Barcelone, Dec. 1999). <http://tecfa.unige.ch/~strasser/barcelone/index.htm>.
- [114] P. Pirolli and R. Rao, *Table lens as a tool for making sense of data*, in *Workshop on Advanced Visual Interfaces, AVI-96* (Gubbio, Italy, 1996). <http://citeseer.nj.nec.com/pirolli96table.html>.
- [115] J.-C. Pomerol, P. Brézillon, and L. Pasquier, 'Operational knowledge representation for practical decision making', *Journal of Management Information Systems* **18** (2002), no. 4, 101–116. <http://www-poleia.lip6.fr/~brezil/Pages2/Publications/PBP2002.pdf>.
- [116] B. Preiss, R. Black, C. Caron, and D. Shapcott, 'Graphic summaries of expert knowledge for the medical curriculum : an experiment in second-year nephrology', *Methods of information in medicine* **31** (1992), 303–310.
- [117] B. Preiss, V. Échavé, S. Preiss, and M. Kaltenbach, 'UVAL-MED a universal visual associative language for medicine', in *proceedings of the 18th annual symposium on computer applications in medical care* (1994), 262–266.
- [118] B. Preiss, M. Kaltenbach, J. Zanazaka, and V. Echave, 'Concept graphics : a language for medical knowledge', in *Proceedings of the 16th annual symposium on computer applications in medical care*, Ed. M. Firsse (McGraw-Hill, Baltimore, USA, 8 Nov. 1992), 515–519.
- [119] I. Purves, B. Sugden, N. Booth, and M. Sowerby, *The PRODIGY project - the interactive development of the release one model*, in *Proceedings of the AMIA annual Symposium* (1999). <http://www.amia.org/pubs/symposia/D005532.PDF>.
- [120] Z. Pylyshyn, 'Computing in cognitive science', in *Foundations of Cognitive Science*, Ed. M. I. Posner, pp. 51–91 (The MIT Press, Cambridge, Massachusetts, 1989). <http://citeseer.ist.psu.edu/pylyshyn89computing.html>.
- [121] _____, 'Some primitive mechanisms of spatial attention', *Cognition* (1994), no. 50, 363–384. <http://citeseer.ist.psu.edu/pylyshyn94some.html>.
- [122] Z. Pylyshyn, 'Is vision continuous with cognition? the case for cognitive impenetrability of visual perception', *Behavioral and Brain Sciences* **22** (1999), no. 3, 341–423. <http://citeseer.ist.psu.edu/149079.html>.

- [123] ———, ‘Visual indexes, preconceptual objects, and situated vision’, *Cognition* **80** (2001), no. 1-2, 127–158. <http://rucss.rutgers.edu/ftp/pub/papers/cognition2001-reprint.pdf>.
- [124] R Development Core Team, *R : A language and environment for statistical computing* (Vienna, Austria, 2004). <http://www.R-project.org>.
- [125] S. Rao and K. Card, *The table lens : Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information*, in *Proceedings of CHI’94, ACM* (Apr. 1994). <http://www.ramanarao.com/papers/tablelens-chi94.pdf>.
- [126] A. Rassinoux, R. Miller, R. Baud, and J. Scherrer, ‘Compositional and enumerative designs for medical language representation.’, *Proc AMIA Annu Fall Symp* (1997), 620–624 (eng).
- [127] ———, ‘Modeling concepts in medicine for medical language understanding.’, *Methods Inf Med* **37** (1998), no. 4-5, 361–372 (eng).
- [128] V. Renault, *Organisation de sociétés d’agents pour la visualisation d’informations dynamiques* (Ph.D. thesis, Université de Paris 6, Dec. 2001). <http://www.lip6.fr/reports/lip6.2002.021.html>.
- [129] V. Rialle, N. Noury, L. Bajolle, J.-B. Lamy, G. Virone, F. Duchêne, N. Moha, and J. Demongeot, ‘Le concept d’Habitat Intelligent pour la Santé : considérations techniques et scientifiques pour un service médico-social’, *Revue de Gériatrie* (2003).
- [130] G. Robertson and J. Mackinlay, ‘The document lens’, in *Proceedings of UIST ’93* (Atlanta, Georgia, 1993), 101–108. <http://www2.parc.com/istl/projects/uir/pubs/items/>.
- [131] G. Robertson, J. Mackinlay, and S. Card, ‘Cone trees : animated 3D visualizations of hierarchical information’, in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI ’91)* (2001), 189–194. <http://www2.parc.com/istl/projects/uir/publications/project>.
- [132] M. Sarkar and M. Brown, ‘Graphical fisheye views of graphs’, in *Proceedings of the human factors in computing systems, CHI’92 conference : striking a balance*, Eds. P. Bauersfeld, J. Bennett, and G. Lynch (ACM Press, Mai 1992), 83–91. <http://citeseer.nj.nec.com/sarkar92graphical.html>.
- [133] B. Sattler, B. Lippy, and Tyrone G. Jordan, ‘Hazard Communication : A Review of the Science Underpinning the Art of Communication for Health and Safety’, *Tech. Report* (Occupational Safety & Health Administration (OSHA), 23 May 1997). <http://www.osha-slc.gov/SLTC/hazardcommunications/hc2inf2.html>.
- [134] H. Schmid, *Probabilistic part-of-speech tagging using decision trees*, in *International conference on new methods in natural language processing* (Manchester, UK, 1994).
- [135] B. Seroussi, J. Bouaud, and E. Antoine, ‘ONCODOC : a successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer.’, *Artif Intell Med* **22** (2001), no. 1, 43–64 (eng). [http://linkinghub.elsevier.com/retrieve/pii/S0933-3657\(00\)00099-3](http://linkinghub.elsevier.com/retrieve/pii/S0933-3657(00)00099-3).
- [136] B. Seroussi, J. Bouaud, H. Dreau, H. Falcoff, C. Riou, M. Joubert, C. Simon, G. Simon, and A. Venot, ‘ASTI : a guideline-based drug-ordering system for primary care.’, *Medinfo* **10** (2001), no. Pt 1, 528–32 (eng).
- [137] A. Seyfang, S. Miksch, and M. Marcos, ‘Combining diagnosis and treatment using Asbru’, in *MedInfo*, **10**, no. Pt1 (2001), 533–537. http://www.ifs.tuwien.ac.at/~silvia/pub/publications/mik_keml97.pdf.
- [138] G. Sherson, *The Relevance of Semiotics to the Internet : How Web Designers use Metaphors in Web Development* (Master’s thesis, Victoria University of Wellington, 1999). <http://www.ucol.ac.nz/~g.sherson/papers/semiotics.htm>.
- [139] R. Shiffman, B. Karras, A. Agrawal, R. Chen, L. Marenco, and S. Nath, ‘GEM : a proposal for a more comprehensive guideline document model using XML’, *Journal of the American Medical Informatics Association* **7** (2000), no. 5, 488–498. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=79044>.

- [140] B. Shneiderman, 'The eyes have it : A task by data type taxonomy for information visualizations', *Tech. Report UMCP-CSD CS-TR-3665* (College Park, Maryland 20742, U.S.A., 1996). <http://citeseer.nj.nec.com/shneiderman96eyes.html>.
- [141] R. Sojourner and M. Wogalter, 'The influence of pictorials on evaluations of prescription medication instructions', *Drug Information Journal* **31** (1997), 963–972. <http://www.diahome.org/content/abstract/1997/dij914.pdf>.
- [142] M. Spenke, 'Visualization and interactive analysis of blood parameters with infozoom', *Artificial Intelligence in Medicine* **22** (2001), no. 2, 159–172. <http://citeseer.nj.nec.com/spenke01visualization.html>.
- [143] M. Spenke, C. Beilken, and T. Berlage, 'FOCUS : The interactive table for product comparison and selection', in *ACM Symposium on User Interface Software and Technology* (1996), 41–50. <http://citeseer.nj.nec.com/spenke96focus.html>.
- [144] J. Starren and S. Johnson, 'An object-oriented taxonomy of medical data presentations', *J Am Med Inform Assoc* **7** (2000), no. 1, 1–20. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=61451>.
- [145] E. Sundvall, M. Nyström, H. Petersson, and H. Åhlfeldt, 'Interactive visualization and navigation of complex terminology systems, exemplified by SNO-MED CT', *Studies in Health Technology and Informatics* **124** (2006), 851–856. <http://www.imt.liu.se/~erisu/2006/MIE2006SundvallTermViz.pdf>.
- [146] D. Sutton and J. Fox, 'The Syntax and Semantics of the PROforma Guideline Modeling Language', *Journal of the American Medical Informatics Association* **10** (2003), 433–443.
- [147] D. Thouvenin, 'Informations des patients - Recommandations destinées aux médecins', *Tech. Report* (ANAES, Mar. 2000). <http://www.anaes.fr/>.
- [148] S. Tu and M. Musen, 'A flexible approach to guideline modeling', in *Proceeding of the AMIA Symposium* (Washington, D.C., 1999), 420–424.
- [149] University of North Carolina at Chapel Hill Duke University, 'Optimizing patient comprehension through medicine information leaflets', *Tech. Report* (United States Pharmacopeia, 1998). <http://www.uspdqi.org/pubs/other/PatientLeafletStudy.pdf>.
- [150] P. Vaillant, *Interaction entre modalités sémiotiques : de l'icône à la langue* (Ph.D. thesis, Université de Paris 11 (Orsay), 1997). http://www.revue-texto.net/Inedits/Vaillant/Vaillant_Interaction.html.
- [151] ———, *Sémiotique des langages d'icônes* (Honoré Champion (coll. Bibliothèque de grammaire et de linguistique), Paris, 1999).
- [152] F. van Ham, *Visualization of State Transition Graphs* (Master's thesis, Technische Universiteit Eindhoven, Aug. 2002). http://www.win.tue.nl/~vanwijk/fvh_thesis.pdf.
- [153] C. van Rijsbergen, *Information retrieval* (Butterworths, London, 1979). <http://www.dcs.gla.ac.uk/~iain/keith/>.
- [154] A. Venot and C. Duclos, *Methodology for the analysis and representation of the medical information about drugs in the Summary of Product Characteristics (SPC)*, in *Proceedings of AMIA* (1999). <http://www.amia.org/pubs/symposia/D005253.PDF>.
- [155] J. Vigneron, I. Gindre, M. Daouphars, P. Monfort, S. Georget, E. Chenot, V. Noirez, N. Commun, F. Laurelli, A. Perrin, M. Lux, M. Hoffman, and M. Hoffman, 'Stabilis 2 : an international CD-ROM database on stability and compatibility of injectable drugs', *Journal of the European Association of Hospital Pharmacists (EJHP)* (2004), no. 2, 59–60. <http://www.eahponline.org/upload/ejhp/Informatics.pdf>.
- [156] S. Wachter, J. March, N. Syroid, F. Drews, and D. Westenskow, 'A new metaphor to display critical pulmonary events during anesthesia', *Journal of the american medical informatics association* **9** (2002), no. 6suppl, S108–S109. http://www.jamia.org/cgi/reprint/9/6_suppl_1/S108.

- [157] S. Wachter, B. Markewitz, R. Rose, and D. Westenskow, 'Evaluation of a pulmonary graphical display in the medical intensive care unit : an observational study', *Journal of biomedical informatics* **38** (2005), no. 3, 239–243. <http://dx.doi.org/10.1016/j.jbi.2004.11.003>.
- [158] C. Wallgren, *A new metaphor for editing structured documents*, in *Proceedings of XML Europe 99* (Granada, 1999). <http://www.infloom.com/gcaconfs/WEB/granada99/wal.htm>.
- [159] M. Wertheimer, *Source book of Gestalt psychology*, ch. Gestalt theory (Harcourt, Brace and Co, 1924). http://www.a2zpsychology.com/a2z%20guide/gestalt_theory.htm.
- [160] J. Wolfe, *Seeing*, ch. Visual attention, pp. 335–386 (Academic Press, 2nd ed., 2000). <http://search.bwh.harvard.edu/pdf/DeValois%20Chapter%20%20~approx%20p.pdf>.
- [161] C. Wroe, W. Solomon, A. Rector, and J. Rogers, *Dopamine : A tool for visualizing clinical properties of generic drugs*, in *Proceedings of the Fifth Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP)* (2000). <http://citeseer.nj.nec.com/wroe00dopamine.html>.
- [162] C. Wroe, A. Roberts, and A. Rector., *Drug information sources : Linking a clinical drug ontology to product information databases*, in *Conference proceedings Healthcare computing (HC2001) BCS Health Informatics Committee* (2001).
- [163] J. Zhang, V. Patel, and T. Johnson, 'Medical error : is the solution medical or cognitive?', *Journal of the american medical informatics association* **9** (2002), no. 6 suppl 1, 75–77. http://www.jamia.org/cgi/reprint/9/6_suppl_1/S75.
- [164] J. Zhang, V. Patel, T. Johnson, and E. Shortliffe, 'A cognitive taxonomy of medical errors', *Journal of Biomedical Informatics* **37** (2004), no. 3, 193–204. <http://dx.doi.org/10.1016/j.jbi.2004.04.004>.

Annexe A

Publications

Article dans des revues d'audience internationale avec comité de sélection

- Lamy, J.-B., C. Duclos, V. Rialle, A. Venot. Which graphical approaches should be used to represent medical knowledge? *Stud Health Technol Inform*, pages 719-724, 2005.

Article en cours de soumission

- Lamy, J.-B., Duclos C., Bar-Hen, A., Ouvrard, P., Venot A. VCM, a graphical language providing more rapid access to medical knowledge for physicians : Evaluation under controlled conditions using drug Summaries of Product Characteristics. Soumis au JAMIA (Journal of the American Medical Informatics Association).

Communication dans des manifestations d'audience internationale

- Lamy, J.-B., C. Duclos, V. Rialle, A. Venot. Which graphical approaches should be used to represent medical knowledge? *Medical Informatics Europe MIE2005*. Geneva, 30/8/2005.

Communication dans des manifestations d'audience nationale

- Lamy, J.-B., C. Duclos, V. Rialle, A. Venot. Vers une méthode rigoureuse de conception des langages graphiques s'appuyant sur les sciences cognitives. In F. Poulet and B. Le Grand, éditeurs, 4ème atelier visualisation et extraction de connaissances, 6èmes journées francophones d'Extraction et de Gestion des Connaissances (EGC2006). Lille, France, 17/1/2006.
- Lamy, J.-B., C. Duclos, V. Rialle, A. Venot. Classification et analyse critique des approches graphiques appliquées à la visualisation des connaissances médicales. *Journées Francophones d'Informatique Médicale (JFIM2005)*, pages 1-9. Lille, 13/5/2005.
- Lamy, J.-B., C. Duclos, V. Rialle, A. Venot. Quelles approches graphiques pour faciliter la diffusion des connaissances médicales?. 5ème congrès Internet et Pédagogie Médicale (IPM 2004), pages 39-40. Grenoble, 3/12/2004.

Brevet

- Lamy J.-B., Duclos C., Venot A. Graphical representation of medical knowledge, USA patent, ref : 60/799 653, 12/5/2006.

Annexe B

Glossaire

AFSSAPS Agence Française de Sécurité SANitaire des Produits de Santé. Elle a entre autre pour rôle de valider les RCP écrits par les laboratoires pharmaceutiques.

analyseur syntaxique Logiciel permettant d'extraire à partir d'un corpus de texte l'ensemble des syntagmes nominaux.

ATC (*Anatomical Therapeutical Chemical classification*) Il s'agit d'une classification médicale arborescente permettant de coder les médicaments et les classes médicamenteuses.

base de connaissances médicamenteuses Une base de connaissances médicamenteuses est un logiciel permettant d'accéder aux connaissances sur le médicament et en particulier aux monographies. Par exemple en France, le Vidal électronique, la banque Claude Bernard ou Thériaque.

contre-indications Voir section 3.2.

CIM10 (Classification Internationale des Maladies, 10ème version) Il s'agit d'une classification médicale permettant de coder les pathologies, à l'origine écrite pour coder les causes de mortalité. La CIM10 est réalisée par l'OMS (Organisation Mondial pour la Santé).

concept Un concept est une propriété d'un terme. Par exemple, les pathologies sont définies par les concepts "étiologie", "localisation fonctionnelle",..., un traitement médicamenteux par les concepts "voie d'administration", "forme", "pathologie traitée",...

corpus de textes ensemble de textes de même nature, réunis dans le but de leur appliquer des méthodes de traitement automatique du langage naturel (TAL).

effets indésirables Voir section 3.2.

étiqueteur grammatical Logiciel permettant de déterminer la catégorie grammaticale des mots d'une phrase : nom, verbe, adjectif,...

Fisheye (oeil de poisson) Il s'agit d'une technique de base en visualisation d'information. Elle consiste à séparer les informations à visualiser en deux catégories selon le point de vue de l'utilisateur, le *focus* et le *contexte*, et ensuite à afficher plus d'information sur le focus que sur le contexte. Cette technique permet de rendre la visualisation interactive, car l'utilisateur doit positionner le focus. Il existe deux types de *Fisheye* : le *Fisheye* filtrant qui n'affiche que le focus, et le *Fisheye* déformant [132] qui accorde plus de place au focus, tout en affichant le contexte avec moins de détails.

HTML (*HyperText Markup Language*) Langage utilisé pour mettre en forme les pages web sur Internet. Ce langage reprend la grammaire XML.

interactions médicamenteuses Voir section 3.2.

index Un signe associé à l'objet qu'il représente (par exemple de la fumée pour du feu).

langage graphique Un langage permettant d'exprimer des informations ou des connaissances sous forme graphique (icônes, images, pictogrammes,...) et non textuelle. Exemple : les panneaux routiers.

lemmatiseur Logiciel permettant de déterminer la forme canonique des mots : le masculin singulier pour les noms et adjectifs, l’infinitif pour les verbes.

mises en garde Voir section 3.2.

monographie d’un médicament La monographie d’un médicament est un document qui reprend les informations du RCP auxquelles sont ajoutées le prix, le coût du traitement journalier, et l’inscription sur les listes des substances vénéneuses. Les monographies correspondent aux documents contenues dans les bases de connaissances médicamenteuses. Voir section 3.2.

précautions d’emploi Voir section 3.2.

pictogramme Un pictogramme est une image monochrome, comme les pictogrammes des panneaux routiers. Un pictogramme ne comprend normalement que la forme (c’est-à-dire pas de couleur).

RCP (Résumé des Caractéristiques Produits) Un RCP regroupe l’ensemble des connaissances cliniques, pharmaceutiques, pharmacologiques et administratives sur un médicament donné. Voir section 3.2.

sémiotique L’étude des signes et des systèmes de signes, et en particulier des langages graphiques. Voir section 2.1.

signe Un signe graphique dans un langage. D’après C.S. Peirce, il existe trois types de signe : le signe iconique, l’index et le symbole. Ces trois catégories ne sont pas exclusives. Un même signe appartient à l’une ou l’autre selon le sens qui lui est associé : par exemple le dessin d’une imprimante est un signe iconique s’il signifie “imprimante”, un index s’il signifie “impression” et un symbole s’il signifie “enregistrer” (ce dernier cas est à éviter!).

signe iconique Un signe qui ressemble à l’objet qu’il représente. Il peut être compris sans apprentissage préalable, de façon “universel”, en procédant par analogie (voir <http://www.edusud.org/ressources/ntic/ica/index.html>).

symbole Un signe dont la signification est arbitraire et conventionnelle, sans rapport avec l’objet qu’il représente. Les symboles sont nécessaires pour représenter des concepts abstraits.

syntagmes nominaux Nom ou groupe nominaux, par exemple un nom avec les adjectifs qui s’appliquent à ce nom.

TAL (traitement automatique du langage naturel) Opération consistant à traiter de manière automatique et informatisée un corpus de texte en langage naturel (par exemple en français), dans le but d’en extraire de la connaissance (par exemple quels sont les mots qui reviennent le plus souvent ensembles,...).

terme Un terme est un groupe nominal dans un texte (un RCP pour ce qui nous concerne) auquel correspond une entrée dans une classification (classification CIM10 par exemple). Dans un RCP, il peut s’agir d’une pathologie, d’un traitement,...

Thériaque Base de connaissance sur le médicament, indépendante des laboratoires pharmaceutiques. Thériaque est utilisé à l’hôpital mais très peu en médecine de ville.

visualisation d’information L’étude de la représentation graphique d’informations quelles qu’elles soient, et notamment des informations abstraites pour lesquelles il n’y a pas de représentation graphique “évidente”, c’est-à-dire qui n’ont pas de propriétés spatiales/géométriques.

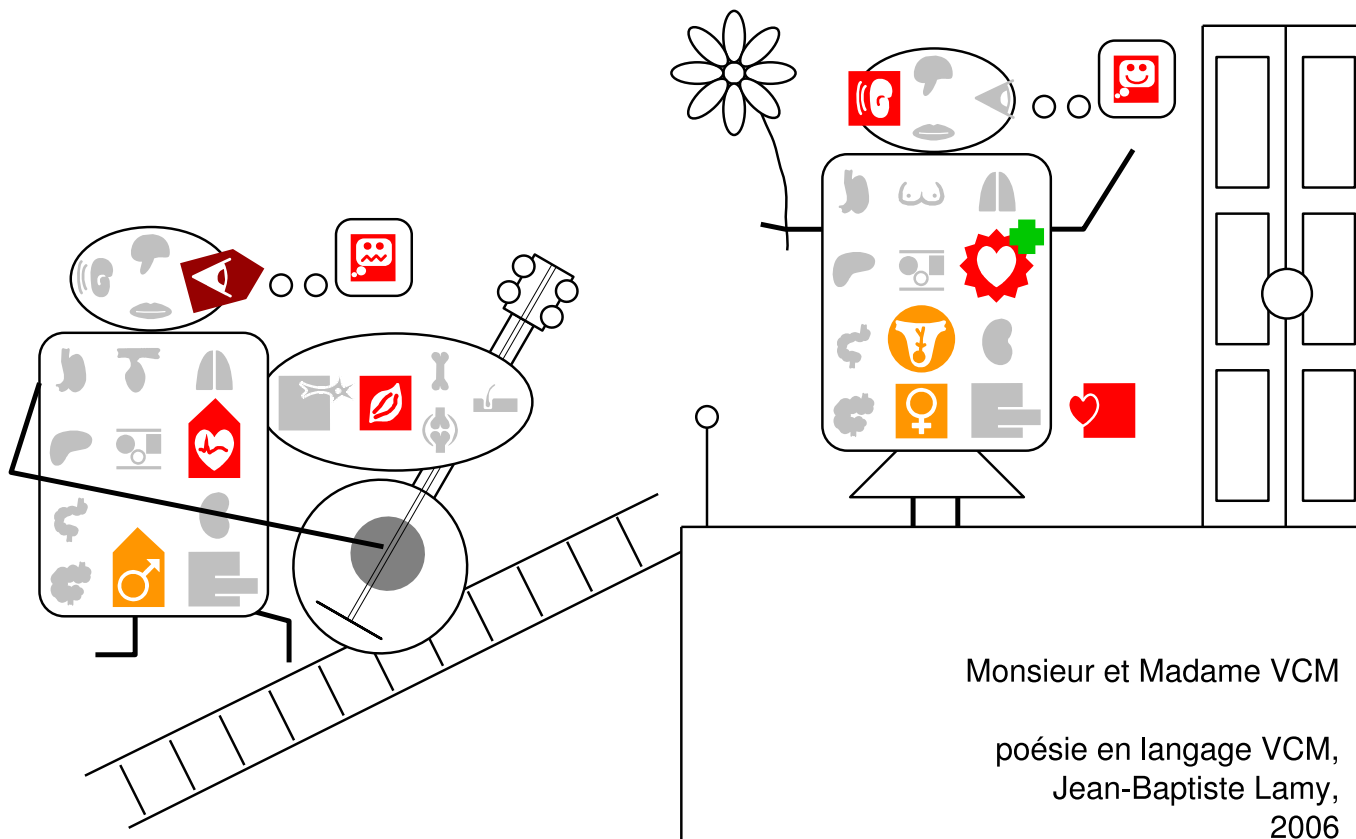
XML (*eXtensible Markup Language*) Grammaire standard permettant d’écrire des “langages de balises”. Un tel langage permet de partir d’un texte, par exemple “Contre-indiqué en cas d’hypertension.” et d’y ajouter des balises, par exemple “Contre-indiqué en cas <terme code=”cim10 :I10”>d’hypertension</terme>.”. Les balises ne sont pas visibles pour l’utilisateur final, cependant les logiciels peuvent en tenir compte ; dans l’exemple ci-dessus, le logiciel “sait” qu’il est question du terme CIM10 I10.

Annexe C

Poésie en langage VCM

Le langage VCM et “Monsieur VCM” ne sont pas dénués d’une certaine esthétique. VCM peut donc donner lieu à des formes d’expression artistique, comme dans l’exemple ci-dessous. Nous parlerons de “poésie” plutôt que de “tableau” ou de “dessin”, puisque l’oeuvre repose sur un langage, et joue avec les règles de ce langage.

Sur le plan artistique, cette démarche qui consiste à s’appuyer un langage graphique médical est très intéressante. En effet, les tableaux ou les photos permettent de montrer les personnages et leurs activités. La bande dessinée, par l’usage des bulles, permet en plus de montrer les paroles ou les pensées conscientes des personnages. Cependant, il n’existe pas encore de moyen pour montrer l’inconscient des personnages. L’utilisation de VCM permet de représenter une partie de l’inconscient, que nous pourrions appeler “l’inconscient physiologique”, c’est-à-dire l’activité de nos organes, présente, passée ou à venir.



Annexe D

Rétrospective de la conception de VCM

La page suivante présente une rétrospective de la conception du langage VCM et de l'évolution de quelques unes des nombreuses variantes qui ont vu le jour.

12/12/2003, «chez les patients hypertendus diabétiques de type 2, ce médicament est indiqué dans le traitement des atteintes rénales».

propofol
> 1 mois

29/1/2004, basé sur les grammaires de Tesnière : «on peut prescrire le propofol chez les patients âgés de plus d'un mois».

16/2/2004, représentation de médicaments sous forme de glyphes ; les couleurs correspondent aux localisations anatomiques (rouge : coeur, bleu : poumon,...).

14/6/2004, inspiré des panneaux routiers : «risque de photosensibilisation».

30/8/2004, inspiré des panneaux routiers : «risque de photosensibilisation».

ARKOGÉLULES MILLEPERTUIS, gélule

9/9/2004, inspiré des graphiques conceptuels.

Rétrospective de l'évolution du langage graphique VCM

14/10/2004, avec des icônes pleines : «CI avec un anti-dépresseur inhibiteur de la sérotonine».

5/1/2005, premier bonhomme (ici les CI et IM du Millepertuis).

5/1/2005, premier bonhomme (ici les CI et IM du Millepertuis).

10/3/2005, panneaux routiers couleur pastel : «chez les patients prenant certains anti-hypertenseurs oculaires, risque de troubles du rythme ; il faut surveiller le rythme cardiaque».

10/5/2005, inspiré des feux rouges : «chez les patients prenant certains anti-fongiques, risque de troubles du rythme ; il faut surveiller le rythme cardiaque».

18/10/2005, la grammaire en version quasi-finale : «CI avec des anti-asthmatiques».

Annexe E

Manuel d'apprentissage du langage VCM

Ce manuel correspond au document papier distribué aux 11 médecins évaluateurs. Il était accompagné du dictionnaire des icônes d'états physiologiques, de pathologies, de médicaments et de surveillances (voir annexe F) et du lexique des pictogrammes (voir annexe G), ainsi que d'un didacticiel sur CD-ROM¹. Le manuel d'apprentissage, accompagné du dictionnaire et du lexique, est aussi disponible sur internet en français : http://vcm.limbio-paris13.org/guide_vcm/vcm_fr.pdf et en anglais : http://vcm.limbio-paris13.org/guide_vcm/vcm_en.pdf.

E.1 Introduction

“VCM” est un langage de Visualisation des Connaissances Médicales destiné aux professionnels de santé. VCM permet de représenter sous forme d'icônes et de pictogrammes des connaissances médicales. Les icônes permettent de repérer les passages qui vous intéressent dans un texte médical ; elles n'ont pas pour objectif de se substituer au texte mais plutôt de servir à “faire le tri” plus rapidement entre les passages qui vous intéressent et ceux qui ne vous intéressent pas pendant la consultation.

La version actuelle de VCM se concentre sur les connaissances sur le médicament, et en particulier sur les RCP (Résumés des Caractéristiques Produit, correspondant aux monographies des médicaments). L'objectif est de vous permettre de trouver plus rapidement les réponses aux questions suivantes :

- “Quelles sont les principales propriétés de ce médicament ?”
- “Puis-je prescrire ce médicament à mon patient ?”
- “Que dois-je dire à mon patient concernant ce médicament ?”

VCM distingue 5 concepts de base, chacun associé à une couleur :

- l'état actuel du patient en **rouge**, qui comprend :
 - les pathologies et les symptômes,
 - les états physiologiques (grossesse, allaitement) et les caractéristiques du patient (âge, sexe, mode de vie,...),
- les risques que court le patient en **orange**,
- les antécédents du patient en **marron**,
- les médicaments en **vert**,
- les surveillances en **bleu**.

Chaque concept est représenté par un assemblage de formes géométriques, de couleurs et de pictogrammes que nous appellerons une icône. Cependant, VCM ne représente que des concepts relativement simples et généraux ; les concepts plus précis sont représentés par le concept plus général qui les englobe.

¹Contactez-moi si vous souhaitez disposer de ce CD-ROM.

Avec ces icônes de concept, VCM permet ensuite de construire des phrases simples pour exprimer les contre-indications, des effets indésirables,... des médicaments.

E.2 L'état actuel du patient

L'état actuel du patient est représenté par la couleur **rouge**. Cette couleur a été choisie pour évoquer la notion de maladie. L'état actuel du patient comprend essentiellement les pathologies et symptômes, mais aussi les états non-pathologiques.

E.2.1 Les états physiologiques et le mode de vie du patient

Une caractéristique du patient est représentée par un rond rouge avec un pictogramme blanc à l'intérieur ; le rond signifie "état normal non pathologique" et le pictogramme représente l'état ou le mode de vie. Par exemple :



Enfant



Consommation d'alcool

E.2.2 Les pathologies et les symptômes

VCM ne différencie pas les pathologies et les symptômes, les pathologies étant fréquemment représentées par leur principal symptôme. "Pathologie" ou "symptôme" sans plus de précision sont représentés par un carré rouge. Le carré signifie "modification", "altération", "pathologie".

Les pathologies et les symptômes d'une localisation anatomico-fonctionnelle sont représentés par un carré rouge avec un pictogramme blanc à l'intérieur qui indique la localisation, par exemple :



Pathologie cardiaque



Alcoolisme

Notons que les pathologies peuvent reprendre les mêmes pictogrammes que les états physiologiques ou les modes de vie, mais dans un carré au lieu d'un rond : l'alcoolisme est une "pathologie" de la consommation d'alcool.

Pour représenter des pathologies plus précisément, VCM distingue deux types de pathologies :

- **Les pathologies qui peuvent s'appliquer à plusieurs localisations**, par exemple, les insuffisances (insuffisance cardiaque, insuffisance rénale, insuffisance hépatique,...), les infections (infection pulmonaire, infection urinaire,...), les cancers et processus tumoraux (cancer du poumon, cancer de l'intestin,...), les hémorragies (hémorragie cérébrale, hémorragie digestive, hémorragie tout court). Dans ce cas, la forme du carré rouge est remplacée par une autre forme qui précise la pathologie, par exemple :



Insuffisance cardiaque

- **Les pathologies qui ne peuvent s'appliquer qu'à une seule localisation**, par exemple les troubles du rythme cardiaque (c'est pourquoi on parle généralement de trouble du rythme sans préciser la localisation). De même, les vomissements sont spécifiques à l'estomac et l'épilepsie au SNC. Dans ce cas, un pictogramme différent est utilisé, par exemple :



Troubles du rythme

Les deux méthodes peuvent être combinées ensemble, par exemple :



Tachycardie (augmentation du rythme cardiaque)

VCM ne dispose de formes et de pictogrammes que pour les pathologies les plus importantes. Les pathologies trop rares ou trop précises sont représentées par une icône plus générale, par exemple nous représenterons “anomalie d’un septum cardiaque” comme :



Pathologie cardiaque sans plus de précision

Les différentes formes utilisées pour exprimer les pathologies sont les suivantes :



Augmentation / hyper-fonctionnement, par exemple :



Hyperthyroïdie (hyper-fonctionnement de la thyroïde)



diminution / hypo-fonctionnement / insuffisance, par exemple :



Insuffisance cardiaque (insuffisance du coeur)



Diminution totale / arrêt, par exemple :



Arrêt cardiaque (arrêt du coeur)



Douleur, par exemple :



Céphalée (douleur du SNC)



Inflammation, par exemple :



Arthrite (inflammation des articulations)



Atteinte nerveuse, par exemple :



Troubles de la conduction (atteinte nerveuse au niveau du coeur)



Atteinte vasculaire, par exemple :



Pathologie coronarienne (atteinte vasculaire au niveau du coeur)



Vaisseau bouché / embolie / thrombose, par exemple :



Infarctus du myocarde (vaisseau bouché au niveau du coeur)



Troubles tensionnels, dont :



Hypertension artérielle (augmentation de la tension)



Hypotension artérielle (diminution de la tension)



Saignement / hémorragie, par exemple :



Hémorragie digestive (hémorragie au niveau de l'intestin)



Oedème, par exemple :



Oedème pulmonaire



Infection, qui se subdivise en :



Infection bactérienne, par exemple :



Méningite bactérienne (infection bactérienne au niveau du SNC),



Infection virale, par exemple :



Hépatite virale (infection virale du foie)



Infection fongique, par exemple :



Infection fongique digestive



Infection parasitaire, par exemple :



Parasitose intestinale



Cancer / tumeur, par exemple :



Cancer du poumon



Allergie / hypersensibilité, par exemple :



Rhinite allergique (allergie au niveau du nez et de la gorge)

E.3 Les risques de pathologies ou de symptômes

Les risques de pathologies ou de symptômes sont représentés exactement comme les pathologies ou les symptômes correspondants, mais en **orange**, par exemple :



Risque de tachycardie

E.4 Les antécédents de pathologies

Les antécédents de pathologies sont représentés exactement comme les pathologies correspondantes, mais en **marron**, par exemple :



Antécédents d'infarctus

E.5 Les traitements

Un traitement est représenté par la couleur **verte**.

E.5.1 Les types de traitement

Les traitements non médicamenteux sont représentés par un carré vert avec un pictogramme blanc décrivant le type de traitement. Les traitements médicamenteux sont représentés par des croix vertes :



Régime diététique



Activité physique



Chirurgie



Traitement médicamenteux

E.5.2 Les médicaments

Un médicament est décrit par la pathologie qu'il traite. L'icône est celle de la pathologie, à laquelle une croix verte est ajoutée "en exposant", par exemple :



Traitement de l'insuffisance cardiaque / cardiotonique

Lorsqu'un cela est pertinent, la voie d'administration peut être précisée en remplaçant la croix verte par un des pictogrammes suivants :



Voie topique



Voie systémique orale



Voie systémique injectable

Cela permet de distinguer certaines classes thérapeutiques comme par exemple :



Anti-asthmatiques locaux et



anti-asthmatiques systémiques



Anti-diabétiques oraux et



insulinothérapie

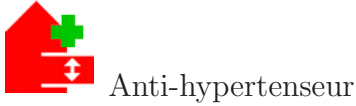


AVK et



héparine

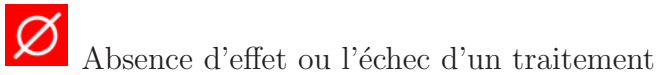
VCM ne dispose de formes et de pictogrammes que pour les principales classes thérapeutiques. Les classes trop rares ou les médicaments indiqués de façon trop précise (par DCI par exemple) sont représentés par une icône plus générale, par exemple nous représenterons le propranolol comme :



Anti-hypertenseur

E.5.3 Les propriétés des traitements

Les propriétés des traitements sont représentées par un carré vert avec un pictogramme blanc à l'intérieur ; le carré vert peut être remplacé par les formes d'augmentation, de diminution ou d'arrêt. Les événements liés à la prise de médicament (surdosage,...) sont représentés sur un carré rouge.



Absence d'effet ou l'échec d'un traitement



Changer la dose



Dose forte



Dose faible



Diminuer la dose



Augmenter la dose



Sevrage / diminution progressive de la dose



Surdosage



Changer le plan de prise

E.6 Les surveillances

Une surveillance est représentée par la couleur **bleu**. Une surveillance est définie à l'aide de deux attributs : le risque surveillé, représenté par une icône de risque (voir ci-dessus), et le type d'examen pratiqué, représenté par un pictogramme bleu parmi les suivants :



Examen clinique



Examen fonctionnel



Examen biologique



Examen d'imagerie

Par exemple :



E.7 Construction de phrases

Dans VCM, les phrases suivent toutes le même modèle :




“conditions” contient les conditions nécessaires pour que la suite de la phrase soit vraie, par exemple une pathologie chez le patient. “affirmations” indique les affirmations, c’est-à-dire effets ou propriétés du médicament (effet indésirable, propriétés pharmacocinétiques,...). “actions_a_faire” et “actions_à_ne_pas_faire” indiquent les actions que le médecin doit faire ou au contraire ne pas faire.

Les conditions sont suivies par une flèche, les actions à faire sont entourées et les actions à ne pas faire sont entourées et barrées. Les propriétés et effets du médicament du RCP sont représentés tels quels, ce qui permet de représenter une propriété (effet indésirable, propriété pharmacologique,...) simplement par son icône. Ces quatre éléments sont tous optionnels ; les éléments absents sont cachés. Par exemple, s’il n’y a pas de condition, le bloc “conditions” n’est pas affiché et la flèche qui le suit peut être enlevée.

Lorsque plusieurs phrases VCM sont nécessaires pour traduire une précaution d’emploi, ces phrases sont placées l’une à côté de l’autre et séparées par un point virgule.



Le symbole  (actions à ne pas faire sans icône à l’intérieur) signifie “ne pas prescrire le médicament du RCP” ou “ne pas se placer dans cette situation”. Il est principalement utilisé pour les contre-indications.

E.7.1 Combinaisons de plusieurs icônes en une seule

VCM peut combiner plusieurs icônes ensembles et les regrouper dans une icône plus générale, pour gagner de la place. Par exemple “trouble du rythme” et “angor” peuvent être combinés en “pathologies cardiaques”. Dans ce cas, une ombre est ajoutée à l’icône, pour indiquer qu’il y a plusieurs icônes “empilées” les unes sur les autres, et donc plusieurs pathologies.

Par exemple :



Pathologie cardiaque



Plusieurs pathologies cardiaques (regroupées en une seule icône)

E.7.2 Négations

La négation ou l’absence se représente par une croix, par exemple :



Patient sédentaire (absence d’activité physique)

E.7.3 Actions

Les actions à faire ou à ne pas faire peuvent être de différents types :

Prescrire un traitement : dans ce cas on utilise l’icône du traitement ou du médicament à prescrire, par exemple :



Il est conseillé de prescrire un régime diététique



Il est conseillé de prescrire un médicament anti-émétique

Modifier ou arrêter le traitement : dans ce cas on utilise les icônes correspondantes, par exemple :



Il est conseillé d'augmenter la dose



Il est conseillé de diminuer la dose



Il est conseillé de changer le plan de prise



Il est conseillé d'arrêter le traitement



Il ne faut pas arrêter le traitement

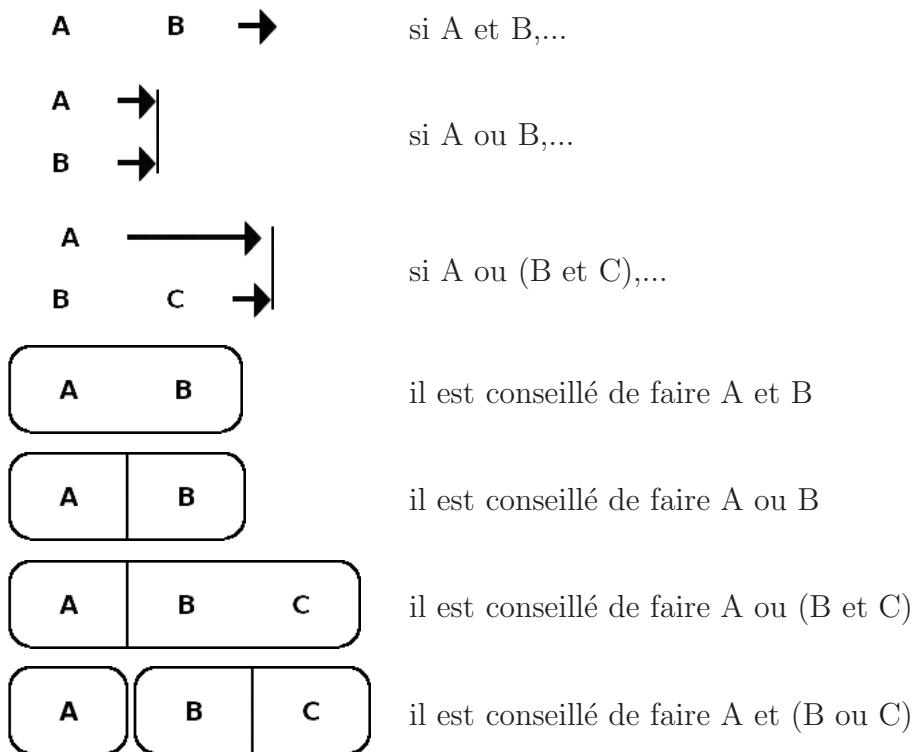
Effectuer une surveillance, par exemple :



Il est conseillé d'effectuer une surveillance électrocardiographique

E.7.4 Relations ET et OU

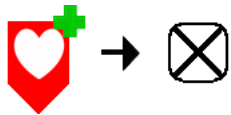
Les relations logiques ET et OU sont représentées différemment dans les conditions et les conseils :



E.7.5 Exemples



Contre-indiqué en cas de paludisme



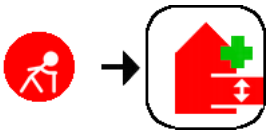
Contre-indiqué avec un médicament de l'insuffisance cardiaque



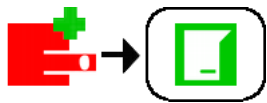
Effets indésirables : troubles digestifs et bradycardie



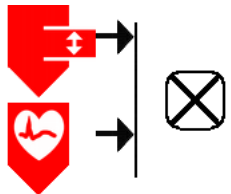
Effectuer une surveillance hépatique



Chez la personne âgée, prescrire un traitement anti-hypertenseur



Précaution d'emploi avec des traitements anti-thrombotiques : prescrire à faible dose



Contre-indiqué en case d'hypotension ou de bradycardie



Contre-indiqué avec l'absorption d'alcool



Chez l'adulte, prescrire des règles hygiéno-diététiques incluant un régime diététique et une activité physique



En cas de forte posologie, effectuer une surveillance clinique ou une surveillance électrocardiographique



Risque d'atteinte hépatique qui nécessite une surveillance de la fonction hépatique ; si un problème hépatique survient, arrêter le traitement.

Annexe F

Dictionnaire VCM-français

Dictionnaire VCM-français des états physiologiques et du mode de vie du patient

VCM distingue quatre classes d'âge :



Nourrisson



Enfant



Adulte



Personne âgée



Sexe masculin



Sexe féminin



Grossesse



Allaitement



Conduite automobile, utilisation de machine



Voyage / origine ethnique



Sportif



Alimentation



Consommation d'alcool



Exposition au soleil

Dictionnaire VCM-français des pathologies et symptômes



Alcoolisme



Tabagisme



Obésité



Infection



Infection virale



Infection bactérienne



Infection fongique



Infection parasitaire



Cancer, tumeur



Cancer du poumon

etc



Maladies de l'appareil digestif



Maladies de la bouche



Maladies de l'oesophage et de l'estomac



Reflux gastro-oesophagien



Nausée / vomissement



Ulcère de l'estomac (estomac + paroi attaquée)



Maladies de l'intestin grêle et du transit intestinal



Infections fongiques digestives (exemple : candidose)



Diarrhée (augmentation du transit intestinal)



Constipation (diminution du transit intestinal)



Occlusion intestinale (arrêt du transit intestinal)



Maladies inflammatoires de l'intestin



Hémorragie digestive



Maladies du colon, du rectum et de l'anus



Maladies du foie



Hépatite virale



Insuffisance hépatique



Maladies de la vésicule biliaire



Lithiase biliaire



Maladies nutritionnelles



Anorexie (diminution de l'alimentation)



Maladies endocrines et métaboliques



Troubles enzymatiques



Maladies de la thyroïde



Hyperthyroïdie



Hypothyroïdie



Diabète



Hypoglycémie



Dyslipidémie



Maladies cardiaques



Pathologie coronarienne (maladie circulatoire au niveau cardiaque)



Infarctus du myocarde (vaisseau bouché au niveau cardiaque)



Troubles du rythme (coeur + signal ECG)



Tachycardie (augmentation du rythme cardiaque)



Bradycardie (diminution du rythme cardiaque)



Troubles de la conduction (trouble nerveux au niveau cardiaque)



Insuffisance cardiaque



Arrêt cardiaque (arrêt du coeur)



Maladies vasculaires



Troubles tensionnels



Hypertension



Hypotension (dont orthostatique)



Maladies du sang



Troubles hydro-électrolytiques



Maladies des globules rouges



Paludisme



Anémie (diminution fonctionnelle et/ou en nombre des globules rouges)



Maladies des plaquettes et de la coagulation



Hémophilie



Maladies des globules blancs et de l'immunité



Leucémie



Déficit immunitaire (dont neutropénie, agranulocytose)



Hémorragie



Thrombose



Maladie de l'appareil respiratoire



Pneumopathies virales



Pneumopathies bactériennes



Cancer du poumon



Embolie pulmonaire



Insuffisance respiratoire



Bronchite



Asthme, BPCO (diminution fonctionnelle des bronches)



Oedème pulmonaire



Maladie du système nerveux



Méningite virale



Méningite bactérienne



Maladies cérébrovasculaires



Hémorragies cérébrales



Migraine, céphalée (douleur du SNC)



Epilepsie (augmentation de l'activité cérébrale)



Maladie de Parkinson (SNC + tremblement)



Maladie d'Alzheimer (SNC + "trou (de mémoire)")



Maladies psychiatrique et comportementales



Maladies de l'humeur



Troubles maniaques (excès d'humeur)



Dépression (diminution de l'humeur)



Troubles du sommeil



Somnolence, hypersomnie (augmentation du sommeil)



Insomnie (diminution du sommeil)



Dépendance médicamenteuse



Vertiges



Anxiété



Asthénie



Maladies des os



Fracture



Ostéoporose (os + trou)



Maladie de Paget (os remodelé)



Maladies des articulations



Goutte (douleur aux articulations)



Arthrite (inflammation articulaire)



Arthrose (diminution fonctionnelle des articulations)



Maladies musculaires



Myalgie



Myasthénie



Maladies ophtalmologiques



Glaucome



Diminution des capacités visuelles (dont myopie, diplopie,...)



Cécité



Atteinte du nerf optique



Maladies de l'oreille et de l'audition



Otite infectieuse



Surdit  (diminution de l'audition)



Maladies du nez et de la gorge



Rhinite infectieuse



Rhinite allergique



Toux



Maladies de la peau



Inflammation cutan e



Photosensibilisation (peau + soleil)



Psoriasis (peau + plaques rouges)



Acn  (peau + boutons)



Rash, prurit,  ryth me,...



Alop cie



Maladies du rein



Cancer du rein



Insuffisance r nale



Insuffisance r nale terminale (arr t du fonctionnement du rein)



Maladies de l'appareil urinaire



Infection urinaire



Cancer de la vessie



Lithiase urinaire



Maladies de l'appareil génital masculin



Maladies de la prostate



Cancer de la prostate



Impuissance (trouble de l'érection)



Maladies de l'appareil génital féminin



Cancer de l'utérus



Cancer de l'ovaire



Métrorragie



Maladies du sein



Cancer du sein



Gynécomastie



Troubles de la lactation

Dictionnaire VCM-français des traitements



Anti-infectieux



Anti-viraux



Antibiotiques




Anti-mycosiques




Anti-parasitaires



Anti-néoplasiques

 Traitements de pathologies immunitaires (dont immuno-suppresseurs)

 Immuno-stimulants

 Analgésiques

 Anti-inflammatoires

 Traitements digestifs

 Traitements de l'estomac

 Anti-acides, traitements des maux d'estomac

 Anti-ulcèreux

 Anti-émétiques

 Traitements intestinaux


 Anti-infectieux intestinaux

 Anti-diarrhéiques


 Laxatifs

 Anti-diabétiques

 Insuline

 Anti-diabétiques oraux

 Vitamines, compléments alimentaires

 Traitements cardio-vasculaires



Cardiotoniques



Anti-angoreux



Anti-arythmiques



Hypolipémiants



Anti-hypertenseurs



Traitements en hématologie



Anti-thrombotiques



AVK



Héparine et dérivés



Anti-hémorragiques



Anti-anémiques



Anti-paludéens



Traitements en pneumologie



Anti-tussifs



Traitements des allergies pulmonaires (anti-histaminiques et corticoïdes)




Anti-asthmatiques



Anti-asthmatiques topiques



Anti-asthmatiques oraux systémiques

 Traitements neurologiques


 Anti-migraineux


 Anti-épileptiques

 Anti-Parkinsoniens

 Traitements psychiatriques

 Psycholeptiques

 Anxiolytiques

 Sédatifs et hypnotiques

 Psychoanaleptiques

 Anti-dépresseurs

 Traitements osseux

 Traitements articulaires

 Anti-gouteux

 Traitements musculaires

 Myorelaxants

 Traitements en ophtalmologie

 Anti-infectieux locaux

 Anti-glaucomeux



Traitements otologiques



Anti-infectieux locaux



Traitements dermatologiques



Anti-septiques, désinfectants



Anti-fongiques dermatologiques



Anti-psoriasis



Anti-acnéiques



Traitements urologiques



Traitements des calculs urinaires



Traitements des troubles de l'érection



Traitements de la prostate



Traitements gynécologiques



Anti-infectieux et anti-septiques gynécologiques



Contraceptifs oraux (traitement du risque de grossesse)

Concepts liés aux médicaments



Dose



Dose forte



Dose faible



Diminuer la dose



Augmenter la dose



Sevrage / diminution progressive de la dose



Plan de prise



Forme thérapeutique



Arrêt du traitement



Présence d'un excipient à effet notable



Absence d'effet ou l'échec d'un traitement



Surdosage



Fausse certains résultats d'analyse

Dictionnaire VCM-français des surveillances



Surveillance clinique



Surveillance de la tension artérielle



Surveillance fonctionnelle



Surveillance pulmonaire fonctionnel



ECG



EEG



Surveillance biologique



Surveillance rénale (clairance)



Surveillance hépatique (transaminases)



Surveillance thyroïdienne



Surveillance de la coagulation (INR, Taux de prothrombine)



Surveillance par imagerie



Surveillance radiologique pulmonaire

















Surveillance radiologique des os

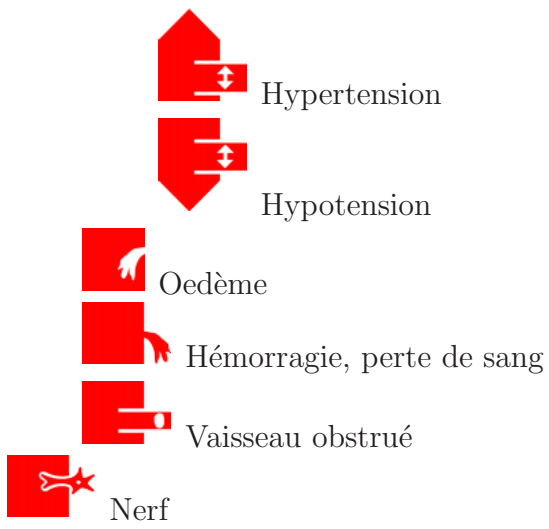
Annexe G

Lexique des pictogrammes du langage VCM

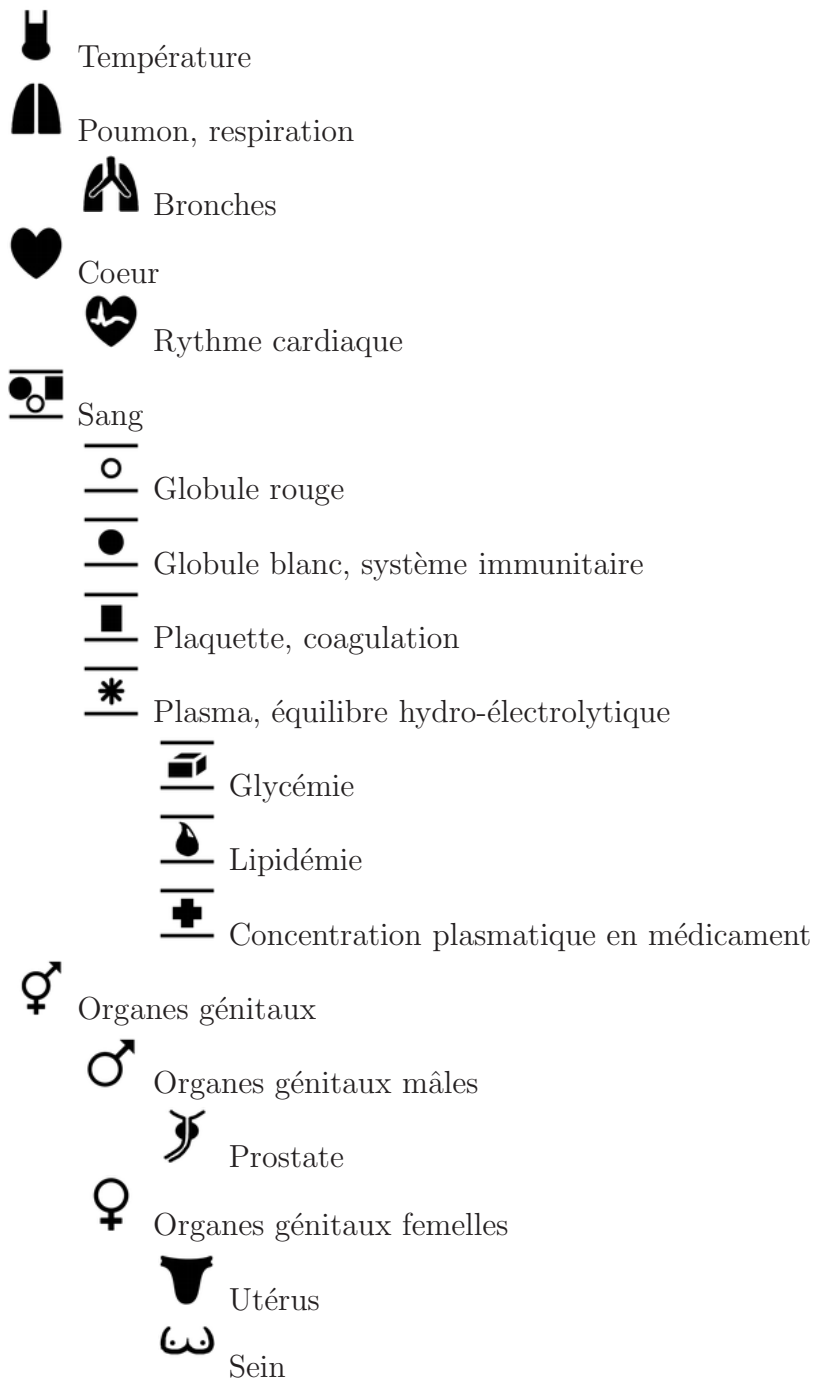
Lexique des pictogrammes et des formes

Formes

-  état normal, non pathologique
-  Hyper-fonctionnement, augmentation, augmenté
-  Hypo-fonctionnement, diminution, diminué
-  Arrêt
-  Inflammation
-  Douleur
-  Infection
-  Infection bactérienne
-  Infection virale
-  Infection fongique
-  Infection parasitaire
-  Processus tumoral
-  Réaction allergique
-  Vasculaire
- Pression artérielle



Organes, systèmes et fonctions



👶 Allaitement



Grossesse



1er trimestre de grossesse



2ème trimestre de grossesse



3ème trimestre de grossesse



Rein, système urinaire



Vessie, voie urinaire



Lithiase urinaire



Alimentation, nutrition



Bouche



Gorge, nez



Toux



Tube digestif, digestion



Estomac



Nausée, vomissement



Reflux gastro-oesophagien



Ulcère



Intestin grêle



Gros intestin, colon, rectum



Foie



Vésicule et voies biliaires



Lithiase de la vésicule biliaire



Autres glandes, système endocrine



Thyroïde



Troubles enzymatiques



Diabète




Cerveau, système nerveux central




Activité cérébrale



Maladie d'Alzheimer

 Maladie de Parkinson, syndrome Parkinsonien

 Comportement, fonctions cérébrales

 Sommeil

 Humeur

 Asthénie, fatigue

 Anxiété

 Vertige

 Pharmaco-dépendance

 Oeil, vision


 Glaucome

 Oreille, audition, équilibre


 Peau

 Photosensibilisation

 Psoriasis

 Rash cutané, prurit, érythème

 Acné

 Alopécie, chute des poils


 Os

 Fracture osseuse


 Ostéoporose

 Maladie de Paget


 Articulation

 Muscle, tissus conjonctifs

Caractéristiques du patient

 Âge : nourisson











 Âge : enfant

 Âge : adulte




 Âge : personne âgée

-  Exposition au soleil
-  Conduite automobile
-  Géographie
-  Sportif
-  Alimentation
-  Alcool
-  Tabac
-  Surpoids, obésité




Traitements

-  Arrêt du traitement
-  Traitement médicamenteux, monothérapie
-  Chirurgie
-  Anesthésie générale
-  Dose, posologie
 -  Dose faible
 -  Dose forte
 -  Surdosage
 -  Sevrage, diminution progressive de la dose
-  Plan de prise, chronopharmacologie

Forme pharmaceutique / voie

-  Comprimé, forme orale
-  Pommade, topique
-  Injectable

Examen

-  Examen clinique
-  Examen biologique
-  Examen d'imagerie

Total : 120 pictogrammes et formes.

Annexe H

Description formelle du langage VCM en BNF

Pour décrire de manière formelle la grammaire de VCM, nous avons utilisé les Formes de Backus-Naur (BNF, http://fr.wikipedia.org/wiki/Forme_de_Backus-Naur) en adaptant un peu ce formalisme. La description en BNF comprend en fait la description de deux langages. Le premier est une représentation en langage quasi-naturel des connaissances médicales représentées. Le second est une représentation en langage quasi-naturel de la grammaire des icônes et phrases VCM. Enfin, une troisième partie fait le lien entre ces deux langages.

Rappel de la syntaxe BNF :

- “:=” signifie “est défini par”
- “<A>” signifie “élément non terminal A”
- “[B]” signifie “éventuellement B”
- “{C}” signifie “un ou plusieurs C”

Éléments de syntaxe supplémentaires :

- “:-” signifie “est représenté (graphiquement) par”
- “{A B ...}” signifie “un ou plusieurs A séparés par des B”, par exemple “A” ou “A B A B A”
- “etc” signifie “liste incomplète”

H.1 Définition des connaissances exprimées par VCM

```
<une phrase> ::= {<une phrase simple>; ...}
<une phrase simple> ::= [<des conditions>] [<des actions à ne pas faire>]
                        [<des affirmations>] [<des actions à faire>]
<des conditions> ::= si {{{<un état du patient> et ...} ou ...}, alors
<des affirmations> ::= {<un état du patient> et ...}
<des actions à faire> ::= il faut faire {{{<une action> et ...} ou ...} et ...}
<des actions à ne pas faire> ::= il ne faut pas faire {<une action> et ...}
<une action> ::= <une prescription d'une surveillance> |
                <une prescription d'un traitement> |
                <une modification d'un traitement>

<un état du patient> ::= <un état du patient non pris en charge> |
                        <un état du patient pris en charge>
<un état du patient non pris en charge> ::= le patient a <un type d'état du patient> <état>
                                           d'<un trouble> [d'<un attribut du patient>]
<un état du patient pris en charge> ::= <un état du patient non pris en charge> pris
                                         en charge par <une prise en charge médicale>
<une prise en charge d'un état du patient> ::= <une prise en charge médicale> prenant en charge
                                                <un état du patient non pris en charge>

<une modification du traitement> ::= <une variation> de <un attribut du traitement>
                                     d'un traitement médicamenteux
<un attribut du traitement> ::= la DCI | la dose | l'arrêt du traitement | le plan de prise

<un type> ::= <un type d'état du patient> |
              <un type de prise en charge médicale>
<un type d'état du patient> ::= un antécédent | un état actuel | un risque
<un type de prise en charge médicale> ::= un traitement médicamenteux |
                                         une surveillance |
```

```

<état> ::= non-pathologique | pathologique

<un trouble> ::= <un trouble générique> | <un trouble spécifique>
<un trouble générique> ::= <une variation> | <une étiologie> |
    <un symptôme ou processus macroscopique> |
    <un symptôme ou processus microscopique>
<une variation> ::= un changement | une augmentation | une diminution | un arrêt
<une étiologie> ::= une infection | une infection bactérienne | une infection virale |
    une infection fongique | une infection parasitaire | une allergie |
    une tumeur
<un symptôme ou processus macroscopique> ::= une douleur | une inflammation |
    une hémorragie | un oedème
<un symptôme ou processus microscopique> ::= une atteinte vasculaire | une atteinte nerveuse |
    une inflammation | une thrombose |
    un trouble tensionnel |
    une hypertension | une hypotension

<un attribut du patient> ::= <une caractéristique> | <une habitude de vie> | <une localisation>
<une caractéristique> ::= l'âge | le sexe | etc
<une habitude de vie> ::= la consommation d'alcool | la pratique sportive | etc
<une localisation> ::= <un système anatomico-fonctionnel> |
    <une localisation anatomique précise> |
    <une propriété d'un système anatomico-fonctionnel>
<un système anatomico-fonctionnel> ::= le coeur | les poumons | etc
<une localisation anatomique précise> ::= les bronches | etc
<une propriété d'un système anatomico-fonctionnel> ::= le rythme cardiaque | etc
<un trouble spécifique> d'<un système anatomico-fonctionnel> ::= troubles du rythme | toux | etc

<une prise en charge médicale> ::= un traitement médicamenteux par <une voie d'administration> |
    une surveillance par <un examen>
<une voie d'administration> ::= voie d'administration non précisée | voie topique |
    voie orale | voie parentérale
<un examen> ::= un examen clinique | un examen fonctionnel |
    un examen d'imagerie | un examen biologique

```

H.2 Définitions des éléments graphiques utilisés par VCM

```

<une phrase graphique> ::= la juxtaposition horizontale de :
    {<une phrase graphique simple> , un point virgule , ...}
<une phrase graphique simple> ::= [<des icônes en ligne> , une flèche ,]
    [<des icônes> entourées et barrées,]
    [<des icônes> sans décoration,]
    {[<des icônes séparées par des barres verticales> entourées]}

<un ensemble d'icônes> ::= <des icônes> | <des icônes en ligne> |
    <des icônes séparées par
    des barres verticales>
<des icônes> ::= {<une icône> , ...}
<des icônes en ligne> ::= {<des icônes> , retour à la ligne, ...}
<des icônes séparées par des barres verticales> ::= {<des icônes> , une barre verticale , ...}]

<une icône> ::= <une icône simple> |
    <une icône avec une icône en exposant>
<une icône simple> ::= <une forme> de couleur <une couleur de forme>
    [avec à l'intérieur <un pictogramme central>]
    {avec autour <un pictogramme additionnel>}
<une icône avec une icône en exposant> ::= <une icône simple> et <une icône en exposant>
<une icône en exposant> ::= en haut à droite <un pictogramme en exposant>
    de couleur <une couleur de pictogramme en exposant>
<une couleur> ::= <une couleur de la forme> |
    <une couleur du pictogramme en exposant>

```

H.3 Correspondance connaissances $\langle \Rightarrow \rangle$ éléments graphiques en VCM

Le membre de gauche contient des connaissances qui ont été définies dans le premier BNF, le membre de droite contient des éléments graphiques qui ont été définis dans le second BNF.

¹Par exemple, celui du RCP.

```

<une phrase>                ::- <une phrase graphique>
<une phrase simple>         ::- <une phrase graphique simple>
<des conditions>           ::- <un ensemble d'icônes> , une flèche vers la droite
<affirmations>             ::- <un ensemble d'icônes> sans décoration
<actions à faire>          ::- <un ensemble d'icônes> entourées
<actions à ne pas faire>   ::- <un ensemble d'icônes> entourées et barrées
ou                           ::- retour à la ligne | barre verticale
et                           ::- ,2

<un état du patient>       ::- <une icône>
<un état du patient non pris en charge> ::- <une icône simple>
<une prise en charge médicale> ::- <une icône en exposant>
<un état du patient pris en charge> ::- <une icône avec une icône en exposant>
<une prise en charge d'un état du patient> ::- <une icône avec une icône en exposant>
<une modification du traitement> ::- <une icône simple>
ce médicament               ::- une case vide
ne pas prescrire ce médicament ::- une case vide entourée et barrée

<un type>                  ::- <une couleur>
un antécédent               ::- marron
un état actuel              ::- rouge
un risque                   ::- orange
un traitement               ::- vert
une surveillance            ::- bleu

<un trouble générique>     ::- <une forme> | <un pictogramme additionnel>
non-pathologique           ::- une forme cercle
pathologique                ::- une forme carré
un changement              ::- une forme carré
une augmentation           ::- une forme carré avec flèche vers le haut
une diminution            ::- une forme carré avec flèche vers le bas
un arrêt                   ::- une forme carré avec flèche vers le bas et trait horizontal
une infection               ::- un pictogramme additionnel qui entre dans le carré
une infection bactérienne  ::- un pictogramme additionnel bactérie qui entre dans le carré
une infection virale        ::- un pictogramme additionnel virus qui entre dans le carré
une infection fongique      ::- un pictogramme additionnel champignon qui entre dans le carré
une infection parasitaire  ::- un pictogramme additionnel ver qui entre dans le carré
une allergie                ::- un pictogramme additionnel molécule qui entre dans le carré
une tumeur                 ::- un pictogramme additionnel deux cellules en division
une douleur                ::- une forme carré avec une explosion
une atteinte nerveuse      ::- un pictogramme additionnel un nerf qui sort du carré
une atteinte vasculaire    ::- un pictogramme additionnel un vaisseau sanguin qui sort du carré
une thrombose              ::- un pictogramme additionnel un vaisseau sanguin bouché
                           qui sort du carré
un trouble tensionnel      ::- un pictogramme additionnel un vaisseau sanguin qui sort du carré
                           avec une flèche dans le vaisseau
une hypertension           ::- un pictogramme additionnel un vaisseau sanguin qui sort du carré
                           avec une flèche dans le vaisseau
une hypotension           ::- un pictogramme additionnel un vaisseau sanguin qui sort du carré
                           avec une flèche dans le vaisseau
une hémorragie            ::- une forme carré avec du liquide qui en sort
un oedème                  ::- une forme carré avec du liquide qui rentre à l'intérieur
une inflammation          ::- une forme carré avec des flammes au dessus

<un attribut du patient>   ::- <un pictogramme central>
<un trouble spécifique> d'<un système anatomico-fonctionnel> ::- <un pictogramme central>
coeur                      ::- un pictogramme coeur
rythme cardiaque          ::- un pictogramme coeur avec ECG
etc

<une prise en charge médicale> ::- <une icône en exposant>
<un type de prise en charge médicale> ::- <une couleur de pictogramme en exposant>
<une voie d'administration> ::- <un pictogramme en exposant>
voie d'administration non précisée ::- pictogramme croix
topique                    ::- pictogramme tube de pommade
orale                      ::- pictogramme comprimé
parentérale                ::- pictogramme seringue
<un examen>                ::- <un pictogramme en exposant>
clinique                   ::- pictogramme stéthoscope
fonctionnelle              ::- pictogramme signal
imagerie                   ::- pictogramme rayonnement
biologie                   ::- pictogramme tube à essai

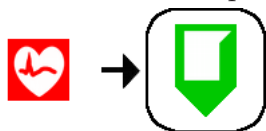
```

²Les relations ET sont représentées simplement en juxtaposant les éléments entre eux. La virgule placée ici ne signifie pas que les relations ET sont représentées par une virgule (auquel cas on aurait écrit "virgule" en toute lettre). Cette virgule est juste un séparateur entre deux éléments à juxtaposer.

H.4 Exemple

Prenons la phrase médicale “En cas de trouble du rythme, diminuer la posologie”. Une fois mise en forme dans le langage décrit par le premier BNF, nous obtenons : “si le patient a un état actuel pathologique de troubles du rythme, alors il faut faire une diminution de la posologie” (ce qui est équivalent à la première phrase, mais plus long car une partie des éléments implicites ont été explicités).

En utilisant le troisième BNF, on peut en déduire que cette phrase sera représentée par “la juxtaposition horizontale de : une forme carrée de couleur rouge avec à l’intérieur un pictogramme coeur avec ECG, une flèche vers la droite, une forme carré avec flèche vers le bas de couleur vert avec à l’intérieur un pictogramme gobelet doseur entourée”. Ce qui est la description en français de la phrase VCM correspondante, que voici :



Annexe I

Classification NPNM

La classification NPNM (Non-Pathologie Non-Médicament) a été développée spécialement pour encoder les connaissances médicales en vue de leur traduction automatique en VCM. Elle comprend l'ensemble des termes qui n'ont pu trouver de traduction satisfaisante ni dans la CIM10 ni dans l'ATC. Il s'agit d'une classification arborescente un peu particulière, puisque chaque terme est soit un fils (héritage), soit un attribut du terme parent.

Les relations entre termes (*est-un*, *a-un*) sont utilisées à des fins de raisonnements, par exemple le terme "rare" hérite du terme "fréquent" car, si le langage graphique ne définit pas la manière de représenté un effet indésirable "rare", il est possible de représenter celui-ci comme un effet "fréquent" à défaut. Par contre l'inverse n'est pas possible, car cela pourrait conduire à des erreurs médicales.

Annexe J

Description du langage XML utilisé pour coder les connaissances médicales

J.1 Présentation

Le langage utilisé pour coder les RCP est un langage XML très simple qui permet d'encoder les connaissances cliniques incluses dans les RCP (ou dans d'autres textes médicaux similaires), avec un niveau de précision assez faible, mais suffisant pour un langage graphique comme VCM. Le prototype que j'ai conçu utilise ces fichiers XML, à la fois pour la représentation "texte + icône" et pour "Monsieur VCM".

Ce langage s'appuie sur les classifications médicales existantes : CIM10, ATC, CDF (classification interne de Thériaque), ainsi que sur une classification NPNM (Non-Pathologie Non-Médicament, c'est à dire ce qui ne rentre ni dans la CIM10 ni dans l'ATC). Lorsqu'un attribut XML contient des codes appartenant à une classification, ils sont formatés de la manière suivante : "<classification>:<code>|<classification>:<code>|...".

Le balisage se fait sur 4 niveaux :

1. le document : il est marqué par la balise <document>, <rcp> ou <theriaque>. Il s'agit de la balise-racine de l'arbre XML. Lorsque cela est pertinent, l'attribut `code` de ces balises indique le code du terme dont traite le document : code ATC du médicament d'un RCP, code CIM10 de la pathologie d'un GBP.
2. les phrases : elles représentent l'unité de sens et correspondent à des phrases VCM. Elles sont marquées par la balise <phrase>, et l'attribut `mod` contient un ou plusieurs code NPNM indiquant le type de phrase (contre-indication,...) ou ses attributs (niveau de contre-indication,...).
3. les relations logiques : elles comprennent les ET et les OU logiques. Elles sont marquées par les balises <et> et <ou>.
4. les termes médicaux : ils sont marqués par la balise <terme>; l'attribut `code` indique le ou les termes, et l'attribut `mod` indique codes des modificateurs NPNM portant sur le terme. Parmi ces modificateurs, nous trouvons les différentes catégories grammaticales (conditions, affirmations, action à faire, action à ne pas faire), la négation, le risque, l'état antérieur (antécédent),... Lorsqu'une balise <terme> comprend un attribut `mod` mais pas de `code`, les modificateurs sont automatiquement reportés sur d'autres termes :
 - la balise parente, si le <terme> est inclus à l'intérieur d'un autre <terme> (<terme code="cim10:I64">un AVC <terme mod="nprm:historique">antérieur<terme></terme>).
 - les balises filles, si le <terme> en contient d'autres (<terme mode="nprm:historique">antécédent d'<terme code="cim10:I64">AVC </terme></terme>).
 - les balises soeurs, en remontant dans la hiérarchie jusqu'à trouver au moins une balise soeur (<terme mode="nprm:historique">antécédent</terme> d'<terme


```
code="cim10:I64">AVC</terme> ou d'<terme code="cim10:i21">infarctus
</terme>).
```

J.2 Exemple

```
<document>

<phrase mod="npm:phrase_contre_indication">
  ce médicament est contre-indiqué
  <et>
    en cas d' <terme code="cim10:I10" mod="npm:condition">hypertension</terme>
    et d' <terme code="cim10:I21" mod="npm:condition|npm:historique">
      antécédents d'infarctus</terme>
  </et>.
</phrase>

</document>
```

J.3 DTD

```
<?xml version="1.0" encoding="UTF-8"?>

<!ENTITY % codes "CDATA"><!-- une liste de code, de la forme :
      "classification:code|classification:code|..."
      avec classification = "cim10", "atc", "npm" ou "cdf" -->

<!ENTITY % balisedetexte "lien">
<!ELEMENT lien      (#PCDATA) >

<!ELEMENT terme    (#PCDATA | terme | %balisedetexte;)* >

<!ENTITY % logical "et | ou">
<!ELEMENT et      (#PCDATA | %logical; | terme | %balisedetexte;)* >
<!ELEMENT ou      (#PCDATA | %logical; | terme | %balisedetexte;)* >

<!ELEMENT phrase   (#PCDATA | %logical; | terme | %balisedetexte;)* >

<!ELEMENT titre    (#PCDATA | terme | %balisedetexte;)* >
<!ELEMENT section  (#PCDATA | titre | section | phrase | terme | %balisedetexte;)* >

<!ELEMENT document (#PCDATA | section | phrase | terme | %balisedetexte;)* >
<!ELEMENT rcp      (#PCDATA | section | phrase | terme | %balisedetexte;)* >
<!ELEMENT theriaque (#PCDATA | section | phrase | terme | %balisedetexte;)* >

<!ATTLIST lien      ref      CDATA   #REQUIRED>

<!ATTLIST terme     code     %codes; #IMPLIED>
<!ATTLIST terme     mod      %codes; #IMPLIED>

<!ATTLIST phrase    mod      %codes; #IMPLIED>

<!ATTLIST section   type     CDATA   #IMPLIED>
<!ATTLIST section   vide     (0 | 1) "0">

<!ATTLIST document  version  CDATA   #FIXED "0.0.1">
<!ATTLIST document  titre    CDATA   #IMPLIED>
<!ATTLIST document  code     %codes; #IMPLIED>

<!ATTLIST rcp       version  CDATA   #FIXED "0.0.1">
```

```
<!ATTLIST rcp      titre  CDATA  #IMPLIED>  
<!ATTLIST rcp      code   %codes; #IMPLIED>
```

```
<!ATTLIST theriaque version CDATA  #FIXED "0.0.1">  
<!ATTLIST theriaque titre  CDATA  #IMPLIED>  
<!ATTLIST theriaque code   %codes; #IMPLIED>
```


Résumé

Les connaissances médicales sont de plus en plus nombreuses et complexes, ce qui rend leur utilisation difficile. Dans d'autres domaines, comme la signalisation routière, des approches graphiques ont permis d'accélérer l'accès aux connaissances. Notre hypothèse est que la médecine pourrait aussi bénéficier de ces approches. Pour cela, nous nous sommes placés dans un cadre multi-disciplinaire : médecine et pharmacie, sciences cognitives et informatique, pour mettre au point des méthodes de visualisation des connaissances médicales, dans le but de faciliter l'accès aux connaissances par les professionnels de santé en situation clinique.

Nous avons suivi une méthodologie rigoureuse s'appuyant sur **(a) la nature des connaissances médicales**, déterminée à partir de la littérature, de connaissances d'experts et d'une analyse par des méthodes de Traitement Automatique du Langage (TAL), et **(b) les capacités de la vision humaine**. Nous proposons **VCM, un langage graphique de Visualisation des Connaissances Médicales** ainsi que "**Monsieur VCM**", un bonhomme graphique interactif. Un prototype appliqué aux connaissances sur le médicament et au Résumé des Caractéristiques Produit (RCP) a été réalisé. Une évaluation dans des conditions contrôlées a montré que **les médecins avaient lu VCM deux fois plus vite que le texte, et en faisant moins d'erreurs**.

Les approches graphiques nous semblent prometteuses pour présenter les connaissances médicales. Les perspectives du langage VCM incluent son extension à l'ensemble de la médecine, l'amélioration de son "universalité" et son utilisation dans d'autres applications médicales.

Mots-clefs : Connaissances médicales, connaissances sur le médicament, résumé des caractéristiques produit, RCP, langage graphique, visualisation d'information, visualisation de connaissance, évaluation

English summary

Medical knowledge is continuously growing and becoming more and more complex, which impairs its clinical use. In other domains, like traffic signs, graphical approaches have allowed a more rapid access to knowledge. Our hypothesis is that medicine can also benefit from these approaches. In this thesis, we choose a multidisciplinary frame : medicine and pharmacy, cognitive sciences, and informatics, for designing methods for visualising medical knowledge, in order to ease the knowledge access by health professionals in clinic situation.

We have followed a rigorous methodology that takes into account **(a) the nature of the medical knowledge**, determined from the literature, from expert knowledge and from an analysis using Natural Language Processing (NLP) tools, and **(b) the human visual abilities**. We propose **VCM, the graphical language for visualising medical knowledge** (*Visualisation des Connaissances Médicales*, in French) and "**Mister VCM**", a graphical interactive schematized body. A prototype applied to drug knowledge and Summary of Product Characteristics (SPCs) has been realized. An evaluation under controlled conditions has shown that **the physicians have read VCM two times faster than text, and have made fewer errors**.

Graphical approaches seems very promising for presenting medical knowledge. The perspectives of the VCM language include the extension of the language to the whole medicine and the improvement of its "universality", as well as its use in other medical application.

Keywords : Medical knowledge, Drug knowledge, summary of product characteristics, SPCs, graphical language, information visualisation, knowledge visualisation, evaluation