



**HAL**  
open science

# Rôle des clients dans la conception d'innovation radicale : le cas du logiciel

Francois Scheid

## ► To cite this version:

Francois Scheid. Rôle des clients dans la conception d'innovation radicale : le cas du logiciel. Humanities and Social Sciences. Ecole Polytechnique X, 2009. English. ⟨NNT : ⟩. ⟨pastel-00005144⟩

**HAL Id: pastel-00005144**

**<https://pastel.hal.science/pastel-00005144v1>**

Submitted on 20 May 2009

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

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



HAL Authorization

Thèse présentée pour obtenir le grade de  
**Docteur de l'Ecole Polytechnique**

Domaine  
**Economie et Sciences Sociales**  
Spécialité  
**Gestion**

Présentée par  
**François SCHEID**

Titre de la thèse :

**Rôle des premiers clients dans la conception  
d'innovation radicale : le cas du logiciel.**

Date de soutenance : le 27 Février 2009

**Composition du jury de soutenance :**

Valérie CHANAL	Professeur des Universités, Université de Grenoble 2 <i>Rapporteur</i>
Mathieu DETCHESSAHAR	Professeur des Universités, Université de Nantes <i>Rapporteur</i>
Florence CHARUE-DUBOC	Chargée de Recherche au CNRS <i>Directeur de Thèse</i>
Robert SALLE	Professeur, EMLYON
Vincent NIBART	TEMIS

A Lancelot, Timothée et Anne-Françoise

A ma grand-mère, Marcelle Blanc.

« L'homme ne se construit qu'en poursuivant ce qui le dépasse. »

André Malraux

## **Remerciements**

Je tiens avant tout à exprimer ma reconnaissance à Florence Charue-Duboc, qui par son exigence, sa disponibilité et sa rigueur a permis de faire converger cette thèse.

Mes remerciements vont aussi à mon épouse, Anne-Françoise et à mes enfants, Lancelot et Timothée, qui ont su faire preuve de patience et de compréhension pour tout le temps que je n'ai pu leur consacrer.

Je remercie aussi mes parents et mon frère pour leur aide précieuse en fin de parcours.

Cette thèse n'aurait pu aboutir sans les « acteurs du terrain », chez Temis et Mondeca en particulier, qui m'ont accueilli avec la plus grande gentillesse et ont su se rendre disponibles malgré le rythme trépidant de leur vie professionnelle. Qu'ils en soient tous remerciés ici.

A EMLYON, je tiens à remercier les membres de l'UPR « Marchés et Innovation », qui ont supporté la tension du chercheur en fin de thèse.

Au CRG, je remercie tous ceux que j'ai rencontrés, doctorants, chercheurs et staff technique, en particulier Michèle, et qui m'ont consacré une partie de leur temps.

# SOMMAIRE

<b>Chapitre introductif</b>	<b>1</b>
<i>Sommaire du Chapitre introductif</i>	2
1. <i>La littérature mobilisée :</i>	4
1.1 Modularité :	5
1.2 L'innovation par les utilisateurs :	7
1.3 Le marketing interorganisationnel :	8
2. <i>L'innovation analysée :</i>	9
3. <i>Méthodologie :</i>	10
4. <i>Plan de thèse :</i>	10
<b>Chapitre 1 : L'industrie du logiciel</b>	<b>12</b>
<i>Sommaire du Chapitre 1 :</i>	13
1. <i>Introduction :</i>	14
2. <i>Une industrie jeune :</i>	15
3. <i>Le logiciel entre produit et services :</i>	18
4. <i>Les difficultés de l'innovation logicielle :</i>	21
5. <i>La création régulière de nouvelles entreprises et de nouvelles applications suivie d'un phénomène de concentration :</i>	22
6. <i>L'architecture logicielle et l'importance des concepts modulaires :</i>	23
7. <i>Stratégie des éditeurs de logiciel et effet de plate-forme :</i>	24
8. <i>Le Web sémantique, et l'émergence des applications de text mining :</i>	36
<b>Chapitre 2 : Apports de la littérature : de la modularisation à l'innovation par les utilisateurs</b>	<b>42</b>
<i>Sommaire du Chapitre 2 :</i>	43
1. <i>La littérature relative à la notion de Modularité :</i>	46
1.1 Introduction :	46
1.2 Définitions :	46
1.3 Les principaux thèmes abordés par la littérature sur la modularité :	50
1.4 La question de l'émergence des <i>design rules</i> :	54
1.5. Deux exemples emblématiques dans l'industrie informatique : les ordinateurs IBM s/360 et le système d'exploitation Unix.	55
1.5.1 Les ordinateurs IBM s/360 :	55
1.5.2 Le système d'exploitation Unix :	57
1.6. Les conséquences de la modularité sur les organisations et la structure d'une industrie :	58
1.6.1 Réseaux de fournisseurs de systèmes modulaires :	58
1.6.2 Architecte ou modulier, deux stratégies exclusives ?	60
1.6.3 Correspondance entre architecture produit et structure organisationnelle :	61
1.7. Modularité et innovation :	62
1.7.1 Introduction :	62
1.7.2 L'innovation relative aux produits modulaires :	63
1.7.3 Architecture produit et exploration des marchés :	65
1.7.4 Limites de l'architecture modulaire :	66
1.7.5 Architecture et processus de conception :	66

1.7.6	Modularité et innovation : conclusion	68
1.8	Conclusion et questions soulevées par la littérature sur la modularité :	68
2.	<i>Les apports de la littérature sur l'innovation par les utilisateurs : le logiciel comme champ d'application idéal ?</i>	70
2.1	Les premiers travaux de von Hippel :	71
2.2	La question de la source de l'innovation : comment les entreprises captent-elles les rentes liées à l'innovation ?	72
2.3	Le concept de <i>lead user</i> :	74
2.4	L'adhérence (« <i>stickiness</i> ») de l'information :	77
2.5	Lien entre la théorie de la diffusion et le concept de <i>lead user</i> :	78
2.6	Le concept de <i>lead user</i> comparé à d'autres concepts de la littérature :	79
2.7	Le concept de <i>lead user</i> dans la littérature :	79
2.8	Une approche « universelle » ?	81
2.9	L'approche <i>lead user</i> pour générer des innovations radicales :	82
2.10	La notion de <i>toolkit</i> : définition et lien avec l'innovation utilisateur	84
2.10.1	Introduction :	84
2.10.3	Le développement des <i>toolkits</i> par le fournisseur :	86
2.11	La contribution des communautés d'utilisateurs :	87
2.12	Conclusion sur l'approche <i>user innovation</i> :	89
3.	<i>La perspective relationnelle retenue par le marketing interorganisationnel :</i>	89
3.1	Introduction :	89
3.2	Les travaux du groupe IMP :	90
3.2.1	La perspective relationnelle :	90
3.2.2	De la dyade au réseau :	91
3.2.3	Le modèle d'interaction :	91
3.2.4	L'innovation comme produit de l'interaction client/fournisseur :	93
3.3	Conclusion :	97
4.	<i>Conclusion sur les apports de la littérature :</i>	98
<b>Chapitre 3 : Méthodologie</b>		<b>101</b>
<i>Sommaire du Chapitre 3 :</i>		102
1.	<i>Les démarches de recherche en gestion et processus d'innovation :</i>	103
1.1	Introduction :	103
1.2	Définition d'un cadre épistémologique et intégrateur :	103
1.3	Les méthodes mises en œuvre dans la recherche sur la conception :	104
1.4	Une approche processuelle :	107
1.5	Conclusion :	108
2.	<i>Le questionnement de l'entreprise :</i>	108
2.1	Projet de recherche :	108
2.2	Questionnement de l'entreprise :	109
2.3.	Construction de la place du chercheur sur le terrain :	111
3.	<i>Les choix méthodologiques :</i>	115
3.1	Une innovation développée au travers de deux projets :	116
3.2	Le projet comme unité d'analyse :	116
3.3	Une analyse processuelle :	117
3.4	L'étude d'une situation emblématique de l'industrie du logiciel :	119
3.4.1	L'étude de la conception d'un nouveau type de logiciel :	119
3.4.2	Les logiciels de « <i>text mining</i> », innovation radicale (ou innovation d'exploration) :	120
3.4.3.	L'accès à une situation emblématique : innovation modulaire multi-acteurs.	121

3.5.	Modalités de l'étude – recueil de données :	122
3.5.1	Modalités de l'étude :	122
3.5.2	Les données recueillies :	124
3.5.3	Les limites de la mise en œuvre de la méthodologie retenue :	128
3.5.4	Conclusion de la partie « Modalités de l'étude - Recueil de données » :	130
4.	Conclusion : une opposition observation/intervention à relativiser.	130
<b>Chapitre 4 : Emergence d'une innovation au travers de deux projets</b>		<b>135</b>
Sommaire du Chapitre 4 :		136
1.	Les éditeurs de logiciel Temis et Mondeca :	139
1.1	Temis :	139
1.2	Mondeca :	141
2.	Les outils de l'annotation sémantique et de l'extraction terminologique :	142
2.1	Le web sémantique :	142
2.2	Constats issus d'un pré-projet en 2002 :	143
2.3	Un outil d'extraction d'information : <i>Insight Discoverer Extractor™</i> de Temis.	146
2.4	Synthèse et ouverture sur l'innovation à développer :	151
3.	Le projet PressPro :	153
3.1	L'origine du projet :	153
3.2	Description et chronologie du projet	154
3.2.1	La solution à mettre en place :	154
3.2.2	La chronologie du projet :	155
3.3	Les phases amont jusqu'à la signature du contrat définitif :	156
3.3.1	Première étape avant la signature d'un premier contrat :	156
3.3.2	Un premier contrat en octobre 2003 : la rédaction des spécifications du projet.	158
3.3.3	Le contrat définitif entre les trois éditeurs signés en mai 2004 :	162
3.4	L'organisation du projet à partir de mai 2004 :	164
3.4.1	Les acteurs du projet :	164
3.4.1.1	Chez Temis :	164
3.4.1.2	Chez Mondeca :	165
3.4.1.3	Chez Xyleme :	166
3.4.1.4	Chez PressPro :	166
3.4.2	Un tissu d'interactions dense :	167
3.4.3	Les étapes du projet :	170
3.4.3.1	Les briques logicielles de Temis :	170
3.4.3.2	L'application documentaire par PressPro, de mai à décembre 2004 :	172
3.4.3.3	Les briques logicielles de Mondeca :	173
3.4.3.4	Le système intégré (à partir de décembre 2004) :	174
3.5	Synthèse –Résumé du projet PressPro :	179
4.	La passerelle logicielle <i>OntoPop</i> et la notion de RAC (règles d'acquisition des connaissances) :	185
4.1	Introduction :	186
4.2	Une passerelle pour l'annotation sémantique et le peuplement d'ontologie :	187
4.3	La formalisation des Règles d'Acquisition de Connaissance	189
4.3.1	L'importance du contexte dans les arbres conceptuels :	189
4.3.2	Méthode et règles d'exploration contextuelle fondement des RAC :	191
4.3.3	Des règles d'exploration contextuelle aux règles d'acquisition de connaissance :	192
4.3.4	Définition des constituants d'une Règle d'Acquisition de Connaissance : PressPro	192
4.3.4.1	Introduction :	192
4.3.4.2	Les composants d'une RAC :	193

4.3.5	Edition des Règles d'Acquisition de Connaissance grâce à un éditeur de règle :	195
4.3.6	Conclusion :	195
5.	<i>Fonctionnement de la passerelle OntoPop :</i>	196
5.1	Cycle vie des RTO dans la démarche OntoPop :	196
5.1.1	L'analyse linguistique :	196
5.1.2	L'application des Règles d'Acquisition de Connaissance :	196
5.1.3	L'enrichissement des lexiques linguistiques de l'outil d'extraction de l'information via le module « mise à jour des lexiques » :	198
5.2	Fonctionnement des modules logiciels d'OntoPop :	198
5.2.1	Le module d'annotation sémantique et de peuplement ontologique :	198
5.2.1.1	La transformation :	199
5.2.1.2	La consolidation :	199
5.2.1.3	La validation :	202
5.2.2	La maintenance des lexiques et autres ressources linguistiques :	202
5.3	Les composants de la passerelle logicielle OntoPop :	203
6.	<i>La méthodologie OntoPop.</i>	204
6.1.	Introduction :	204
6.2	Présentation générale de la méthodologie OntoPop :	205
6.3	Les différentes phases de la méthodologie OntoPop :	208
6.3.1	La Phase d'Etude :	208
6.3.2.	La Phase de Structuration :	213
6.3.2.1	Modélisation de l'ontologie du domaine :	213
6.3.2.2	Construction des cartouches linguistiques :	214
6.3.3	La Phase de Couplage :	218
6.3.4	La Phase de Validation :	220
6.3.5	La Phase de Mise en Service :	222
7	<i>Le projet Exinis :</i>	224
7.1	L'origine du projet Exinis :	224
7.2	Description et chronologie du projet :	225
7.3	Les phases amont jusqu'à la signature du contrat définitif :	228
7.3.1	Premiers contacts en avril 2004 :	228
7.3.2	Un premier contrat pour un démonstrateur en juin 2004 :	228
7.3.3	Présentation du démonstrateur à Exinis US :	229
7.3.4	Le contrat définitif signé en mai 2005 :	230
7.4	Les acteurs et l'organisation du projet (à partir de mai 2005) :	231
7.4.1	Les acteurs du projet chez Exinis :	231
7.4.2	Les acteurs du projet chez Temis :	233
7.4.3	Les acteurs du projet chez Mondeca :	234
7.4.4	Les acteurs du projet chez 4D Concept :	234
7.4.5	L'acteur du projet chez Unilog :	234
7.4.6	Les réunions de projet organisées par l'architecte innovation d'Exinis :	235
7.5.1	Développement de la version V0 de la solution (mai 2005-février 2006):	235
7.5.2	Relance du projet et finalisation de la version V1 (juin 2006 - juin 2008) :	239
7.6	Synthèse - résumé : Principales différences entre les projets PressPro et Exinis.	241
8.	<i>Conclusions et apports des projets PressPro et Exinis :</i>	242
<b>Chapitre 5 : Discussion</b>		<b>246</b>
<i>Sommaire du Chapitre 5 :</i>		247
1.	<i>Introduction :</i>	249

2.	<i>Modularité :</i>	249
2.1	Deux niveaux d'innovation, architectural et local :	249
2.2	L'illusion d'une modularité d'usage :	251
2.3	Les composants de l'innovation logicielle :	252
2.3.1	Des logiciels cœur à la plate-forme de création de base de connaissance :	252
2.3.2	Les composants de la passerelle OntoPop :	254
2.3.3	L'application documentaire :	255
2.3.4	L'émergence d'une nouvelle plate-forme logicielle :	255
2.4	Conclusion sur la nature de l'offre :	256
3.	<i>La structuration des projets: collaboration, coordination et compétences.</i>	258
3.1	Introduction :	258
3.2	Des collaborations interorganisationnelles multiples :	258
3.2.1	Introduction :	258
3.2.2	Les collaborations liées aux différentes tâches :	259
3.2.2.1	La passerelle logicielle OntoPop :	260
3.2.2.2	La cartouche de connaissance personnalisée.	261
3.2.2.3	L'ontologie du domaine étudié :	262
3.2.2.4	La définition personnalisée des RAC :	263
3.2.2.5	L'application documentaire :	263
3.2.2.6	Conclusion :	265
3.3	La nature des interactions entre acteurs :	266
3.3.1	Introduction :	266
3.3.2	Interactions fournisseur/client lors du projet PressPro :	267
3.3.2.1	Interaction des membres du consortium avec l'architecte innovation de PressPro:	267
3.3.2.2	Interaction entre membres du consortium et utilisateurs classiques de PressPro :	269
3.3.2.3	Interaction des membres du consortium avec les informaticiens de PressPro :	271
3.3.2.4	Interaction des membres du consortium avec le management de PressPro :	271
3.3.3	Interaction fournisseur/client lors du projet Exinis :	272
3.3.3.1	Interaction entre Temis, Mondeca et 4D Concept, et l'architecte innovation d'Exinis :	272
3.3.3.2	Interactions entre Temis, Mondeca et 4D Concept, et les utilisateurs d'Exinis :	273
3.3.3.3	Interactions entre Temis, Mondeca et 4D Concept et les informaticiens d'Exinis :	274
3.3.3.4	Interactions entre Temis, Mondeca et 4D Concept et le management d'Exinis :	274
3.3.3.5	Comparaison entre les projets PressPro et Exinis, du point de vue des interactions fournisseur/client :	275
3.3.4	Interaction entre Temis et Mondeca :	275
3.3.5	Conclusion sur la nature des interactions entre acteurs :	277
3.4	Une structuration de projet émergente :	277
3.4.1	Introduction :	277
3.4.2	Structuration du projet PressPro :	279
3.4.3	Structuration du projet Exinis :	282
3.4.4	Une structuration définie après une phase d'exploration :	286
3.4.5	La question de l'encastrement de la coordination :	287
4.	<i>Les rôles joués par le client dans le processus d'innovation : enrichissement de la notion de lead user.</i>	288
4.1	Introduction :	288
4.2	Le rôle du client à plusieurs niveaux :	289
4.2.1	La conception de l'architecture de l'innovation : le premier client à l'initiative du projet.	289
4.2.2	La conception des composants d'interfaçage entre les core products constituant l'innovation :	291
4.2.3	La personnalisation des logiciels core products et de la passerelle OntoPop :	292
4.2.4	La conception de l'application documentaire :	295
4.2.5	La coordination globale du projet :	297
4.3.	Synthèse des rôles joués au cours des projets PressPro et Exinis :	300

4.4	Enrichissement de la notion de <i>lead user</i> :	301
4.4.1	Des caractéristiques de lead users :	301
4.4.2	Nature des rôles joués par les acteurs clefs des projets chez PressPro et Exinis :	302
4.4.3	Enrichissement de la notion de lead user :	303
4.4.3.1	Introduction :	303
4.4.3.2	L'indispensable contribution du client au processus d'innovation :	303
4.4.3.3	Une contribution multiple du client, par des acteurs divers :	304
4.4.4	Conclusion sur la contribution du client au processus d'innovation :	307
4.5	Des tâches qui ne peuvent être assumées par le client seul :	308
5.	<i>Conclusion du chapitre</i> :	309
5.1	La modularité de l'innovation :	309
5.2	L'innovation par les utilisateurs :	310
5.3	L'apprentissage :	314
<b>Conclusion</b>		<b>317</b>
1.	<i>Les apports de la recherche</i> :	318
1.1	Introduction :	318
1.2	Apports dans le champ de la modularité :	319
1.3	Apports dans le champ de l'innovation par les utilisateurs :	323
2.	<i>Limites de la recherche</i> :	325
3.	<i>Poursuite des recherches, pistes de recherches futures</i> :	327
<b>Table des tableaux</b>		<b>328</b>
<b>Table des figures</b>		<b>330</b>
<b>Bibliographie</b>		<b>333</b>
<b>Annexe</b>		<b>352</b>
<i>Sommaire des Annexes</i>		353
<i>Annexes : Web sémantique et notions liées.</i>		354
1.	<i>L'annotation sémantique, définitions</i> :	354
2.	<i>Le problème à résoudre : l'automatisation de la création des annotations en utilisant un moteur d'extraction d'information.</i>	357
2.1.	Création manuelle des patrons : le fonctionnement des moteurs d'extraction d'information (EI).	357
2.2.	Apprentissage supervisé des patrons :	358
3.	<i>Le stockage des annotations et les bases de connaissance</i> :	358
4.	<i>L'utilisation des annotations par les agents logiciels</i> :	359
5.	<i>L'annotation et le Web Sémantique</i> :	359
6.	<i>Les Ressources Terminologiques ou Ontologiques (RTO) : définitions.</i>	359
6.1.	Les Taxonomies :	360
6.2.	Les Thesaurus :	360
6.3.	Les Ontologies	361
6.4.	Les RTO et l'annotation sémantique :	362

# **Chapitre introductif**

## SOMMAIRE DU CHAPITRE INTRODUCTIF

<b>Chapitre introductif</b>	<b>1</b>
<i>Sommaire du Chapitre introductif</i>	2
1. <i>La littérature mobilisée :</i>	4
1.1 Modularité :	5
1.2 L'innovation par les utilisateurs :	7
1.3 Le marketing interorganisationnel :	8
2. <i>L'innovation analysée :</i>	9
3. <i>Méthodologie :</i>	10
4. <i>Plan de thèse :</i>	10

La place du logiciel dans la vie des entreprises, comme dans celle des individus, ne cesse de croître, témoin d'un mouvement puissant vers un « monde numérique » (Berry G., 2008). L'industrie du logiciel se caractérise, quant à elle, par des niveaux de turbulence élevés, qui concernent à la fois les technologies et les marchés. Elle se distingue par un flux d'innovations radicales, qui conduisent à l'émergence régulière de nouveaux types d'application. Ces innovations radicales sont généralement conçues par de petites structures de type start-up technologique, qui consacrent leurs moyens au développement d'un logiciel, ou d'une gamme de logiciels, très spécifiques. Enfin, ces innovations radicales dans le domaine du logiciel souffrent d'un taux d'échec commercial très élevé, qui explique le faible taux de survie de ces start-up (Cusumano, 2004).

Ces différents éléments concourent à éclairer l'intérêt de notre travail de recherche, qui traite de la question de la conception d'innovation radicale dans le domaine des logiciels destinés aux entreprises. Les start-up qui portent des innovations logicielles radicales ont généralement été créées dans le but d'exploiter des travaux de recherche. Parmi les défis auxquels ces entreprises sont confrontées pour leur survie, figurent la capacité à associer le client aux démarches de conception et à coopérer avec lui. La prise en compte des besoins des clients a été soulignée comme un facteur clé de succès des processus d'innovation dès les travaux de Freeman (1982). Von Hippel et Katz (2002) considèrent même que les clients jouent un rôle déterminant dans l'émergence et l'orientation de l'innovation. Conscients du fait qu'un logiciel radicalement innovant n'est qu'un concept tant qu'il n'a pas été implémenté pour une utilisation précise (Cusumano, 2004), les éditeurs de logiciel radicalement innovants sont alors confrontés à la question de savoir comment associer les clients aux premières mises en œuvre du concept innovant.

La vente d'un logiciel radicalement innovant aux premières entreprises clientes passe généralement par une personnalisation du logiciel, visant à l'adapter à l'utilisation précise qui doit en être faite, ainsi qu'au système d'information des utilisateurs (Cusumano, 2004). Une fois le nouveau logiciel conçu par un éditeur, un travail de développement adapté à chacun des premiers clients s'engage alors : il implique l'éditeur de logiciel innovant, un éventuel intégrateur, généralement une société de services informatiques, et les utilisateurs finaux du système selon différentes modalités. Ce travail contribue à la définition de l'innovation elle-même. De même, il

est courant, si ce n'est systématique, que les clients demandent à ce que le logiciel innovant, très spécifique, soit associé à d'autres logiciels, pour parvenir à une solution intégrée : la mise en œuvre d'une telle solution multi-acteurs requiert alors une collaboration plus complexe impliquant le client et plusieurs éditeurs de logiciel.

Notre question de recherche, qui découle de ces constats, peut être formulée de la manière suivante : « quels rôles jouent les premiers clients dans le processus de conception d'innovation logicielle radicale multi-acteurs, et comment les associer à ce processus ? ». Cette question sera étudiée dans le secteur des logiciels destinés aux entreprises. A la rencontre entre des questions théoriques et empiriques, notre recherche porte donc sur le rôle du client et les relations entre plusieurs fournisseurs, dans le développement d'innovations logicielles radicales multi-acteurs. Nous positionnons maintenant la littérature que nous mobiliserons pour analyser ce processus de construction de l'innovation logicielle radicale avec les premiers clients. Nous présentons ensuite l'innovation analysée et le matériau empirique qui nous permettra d'approfondir cette question.

## 1. LA LITTÉRATURE MOBILISÉE :

Pour répondre à notre question de recherche, nous nous appuyerons principalement sur les travaux, qui ne sont pas spécifiques au domaine du logiciel, portant sur le concept de modularité tout d'abord, à celui d'innovation par les utilisateurs ensuite, et enfin nous nous intéresserons aux apports de la recherche en marketing interorganisationnel. Ces choix résultent de différents éléments. Tout d'abord, les concepts de modularité et d'innovation par les utilisateurs sont dominants dans le secteur du logiciel. En effet, depuis les travaux de Parnas (1972), la modularité est un des principes fondamentaux de la conception logicielle. Le découpage d'un logiciel en modules indépendants, facilite l'éclatement du processus de développement entre différents acteurs, éditeurs de logiciel, utilisateurs ou sociétés de services. Ainsi, chacun de ces acteurs peut, à certaines conditions, travailler sur un module particulier, phénomène observé bien sûr dans le champ du logiciel *Open Source*, mais aussi dans celui du logiciel propriétaire, où nous nous situons. Cette contribution est facilitée par l'utilisation de boîte-à-outils ou *toolkit*, permettant à ces acteurs de développer des modules logiciels personnalisés, qui viennent ensuite s'interfacer avec les modules logiciels génériques constitutifs de la solution. On voit ici clairement le lien qui existe entre la modularité des architectures logicielles, et la place prise par les innovations développées par les utilisateurs ou *user innovation* dans ce secteur. Cela explique en partie pourquoi l'innovation par les utilisateurs est un concept dominant dans le logiciel, au point que certains auteurs en

arrivent à affirmer que les utilisateurs constituent la source privilégiée d'innovation dans ce domaine (Franke et von Hippel, 2003). Enfin, nous reviendrons sur les travaux menés en marketing interorganisationnel, qui, à l'instar du concept *d'user innovation*, trouvent leur origine dans la remise en cause d'une approche pour laquelle la conception de l'innovation est intégralement assumée par le fournisseur, en l'occurrence l'approche diffusionniste. Nous retenons ce champ, car le processus d'innovation que nous avons choisi d'étudier passe par la mise en œuvre de collaborations entre plusieurs acteurs, éditeurs de logiciel et clients, aspect clairement pris en compte et souligné par la recherche en marketing interorganisationnel, qui considère l'innovation comme le produit d'un réseau d'acteurs. Le rôle des collaborations entre acteurs dans le processus d'innovation est par contre très peu traité par la littérature sur l'innovation par les utilisateurs. Ainsi, nous essayerons de déterminer si ces approches, apparemment contradictoires, sont susceptibles de s'enrichir l'une l'autre.

## 1.1 Modularité :

La *modularité* est au centre des travaux de recherche sur les processus d'innovation dans le domaine informatique. Dans le logiciel, en particulier, le concept de modularité est systématiquement mis en œuvre depuis les années 1970, comme nous l'avons vu. « Différentes sociétés peuvent concevoir et produire indépendamment des composants tels que des lecteurs de disques durs ou des systèmes d'exploitation, qui sont autant de modules qui fonctionneront ensemble dans un produit complexe, car les concepteurs de ces modules suivent un ensemble de règles de développement, *design rules*. » (Baldwin & Clark, 1997). La *modularisation* est la décomposition d'un système complexe en sous-systèmes quasi autonomes, pouvant être conçus de manière indépendante (Baldwin et Clark, 1997 ; Aoki, 2002). Le problème est alors d'élaborer un système complexe en intégrant ces sous-systèmes. La *modularisation* peut ainsi être vue comme une stratégie de spécialisation et de division du travail, permettant de maîtriser la complexité (Simon 1962, Parnas 1972, Baldwin et Clark 2000, Ethiraj et Levinthal, 2004). La *modularisation* peut également être considérée comme un mode de gestion de l'innovation car elle laisse chaque responsable de module libre d'innover, en respectant les règles édictées par l'architecte. Le concept général de modularité inclut, tout d'abord, deux idées majeures, celle d'interdépendance à l'intérieur d'un module, et celle d'indépendance entre les modules. Ainsi, la modularité est une forme de conception particulière qui crée intentionnellement un haut degré d'indépendance ou « faible couplage » entre des concepts de composant en standardisant les spécifications de l'interface de chaque composant.

## Comment émerge le *dominant design* ?

La littérature sur la modularité traite peu du processus qui aboutit à l'émergence d'un *dominant design* modulaire. La tendance croissante à la *modularisation* des technologies et à la désintégration des systèmes implique la production de nouveaux types de connaissances (Steinmueller, 2002), telles que des normes, des standards, ou encore un savoir d'intégration (Shapiro et Varian, 1999). Ces nouveaux types de connaissance sont nécessaires à la coordination, c'est-à-dire à l'intégration d'ensembles faiblement couplés. Comment ces connaissances émergent-elles ? Qui les construit ? Comment sont élaborées les *design rules* ? Ces questions restent ouvertes.

## Modularité et fonctionnement organisationnel :

Sanchez et Mahoney (1996) suggèrent que si les organisations conçoivent les produits, inversement les produits conçoivent les organisations, car les tâches de coordination qui sont implicites dans les concepts de produit déterminent les concepts d'organisation capables de développer et de réaliser ces produits. Ces auteurs affirment que l'ingénierie traditionnelle suit une méthodologie de conception qui conduit à des concepts de produit, formés de composants fortement couplés, ce qui nécessite une coordination managériale intensive lors de leur développement. L'alternative proposée par la modularisation des architectures produit, consiste à réaliser des interfaces de composant standardisées pour des concepts de composants faiblement couplés. Cela permet alors la coordination effective des processus de développement sans exercice d'autorité managériale (Sanchez et Mahoney, 1996).

Dans le cas des logiciels innovants, nombre d'entreprises sont de taille réduite. Cela va de paire avec un processus de spécialisation (Horn, 1999), et entraîne un morcellement de l'offre qui rend plus complexe la constitution de solutions globales. Une offre constituée de logiciels développés par des sociétés indépendantes les unes des autres, possède une structure qui semble a priori modulaire *par construction*. Nous faisons l'hypothèse que le fait qu'une innovation soit constituée de différents composants logiciels portés par des organisations indépendantes conditionne fortement le processus d'innovation. L'émergence d'une solution modulaire multi-acteurs soulève un certain nombre de questions, que nous nous proposons d'étudier :

Comment émerge la vision globale indispensable à la conception d'une telle solution innovante ? Quelles sont les collaborations nécessaires à la réalisation de l'innovation modulaire multi-acteurs

que nous nous proposons d'étudier ? Quel est le besoin de coordination requis par de telles collaborations ?

## 1.2 L'innovation par les utilisateurs :

Comme annoncé, nous souhaitons articuler la littérature sur la modularité et la littérature relative au concept d'*user innovation*, qui attribue un rôle central aux utilisateurs en conception, pour questionner le rôle qu'ils peuvent assumer dans l'émergence d'une architecture modulaire. Nous présentons rapidement ci-dessous quelques apports majeurs de ces travaux.

L'importance de l'innovation par les utilisateurs ou « *user innovation* », a été mis en exergue par les travaux de von Hippel (1977), qui considère avec d'autres auteurs qu'elle peut permettre la génération de flux continus d'innovations (Von Hippel, Thomke et Sonnack, 1999). Von Hippel (1986) insiste sur le rôle unique que peuvent jouer une catégorie particulière d'utilisateurs, les *lead users*, concept central de ce courant de littérature, dans le processus d'innovation.

### Le logiciel, « produit numérique » : terrain propice à l'innovation par les utilisateurs ?

Von Hippel (1986) suggère ainsi aux entreprises de générer des concepts d'innovation en partenariat avec des *lead users*. La notion de *lead user* est définie comme suit : il s'agit d'utilisateurs (individus ou organisations) qui ressentent des besoins propres qui deviendront ultérieurement ceux d'un grand nombre d'utilisateurs, et qui espèrent retirer des bénéfices significatifs de la satisfaction de ces besoins. De tels utilisateurs ressentent donc une forte motivation à innover, et peuvent imaginer, concevoir, voire développer des innovations qui ensuite connaîtront une grande diffusion. Ainsi, il peut être profitable pour un fournisseur de laisser certains utilisateurs développer eux-mêmes leurs innovations, comme dans le secteur de l'instrumentation scientifique (von Hippel, 1994). C'est pourquoi certaines entreprises ont décidé d'aller au bout de cette logique, en choisissant d'équiper leurs clients d'outils permettant de concevoir et de développer eux-mêmes les produits qui leur conviennent. Von Hippel (2001) propose ainsi une approche dans laquelle l'utilisateur reçoit les outils (« *user toolkit* » ou boîte à outils de l'utilisateur) permettant de finaliser complètement la conception du produit. Le but est ensuite d'incorporer certaines solutions développées par ces utilisateurs dans des produits standards qui intéresseront un nombre plus large de clients (Thomke et von Hippel, 2002). L'approche *toolkit* passe par le découpage du processus d'innovation, en sous-tâches dévolues soit à l'utilisateur soit au

producteur (von Hippel et Katz, 2002), et permet d'envisager une *mass customization* ou personnalisation de masse. Une telle répartition des tâches peut impliquer des changements radicaux quant à l'architecture d'un produit, et entraîne généralement le développement d'une architecture modulaire (von Hippel et von Krogh, 2003). Ainsi favorisée par le caractère modulaire des architectures logicielles, l'approche *toolkit* est répandue dans l'univers du logiciel.

### **Questionnements relatifs à l'innovation par les utilisateurs :**

Le logiciel apparaît clairement comme un terrain d'application privilégié pour l'innovation utilisateur. Ce secteur met en œuvre avec succès les concepts de *lead users* et de *user toolkit*, et ce en grande partie grâce à la modularité de son architecture, mais aussi grâce à son caractère immatériel. Il n'en demeure pas moins que des questions se posent, dans le cas que nous étudions, c'est-à-dire celui de l'innovation radicale multi-acteurs :

Comment se conjuguent ces deux éléments contradictoires, terrain favorable à l'innovation utilisateur d'une part, et difficulté d'appropriation, par le client, des technologies nouvelles qui fondent l'innovation ? Quel rôle peut concrètement jouer un *lead user* dans un tel processus d'innovation ? Est-il décisif ? Quelles compétences doit-il posséder, quelles compétences doit-il développer ?

### **1.3 Le marketing interorganisationnel :**

Fondée, sur la mise en cause d'une approche centrée sur le fournisseur, l'approche diffusionniste, tout comme l'approche *lead user* a été fondée sur la mise en cause du modèle *manufacturer active* dans le processus de conception, les chercheurs en marketing interorganisationnel, étudient les déterminants des transactions entre un fournisseur et son client, en se focalisant sur l'analyse des relations entre ces acteurs sur le long terme, plutôt que sur les transactions elles-mêmes. Chaque relation client/fournisseur englobant de multiples interactions entre acteurs et groupes d'acteurs, l'étude de cette relation requiert la caractérisation de ces différentes interactions. A cette fin, un modèle, appelé « modèle d'interaction » a été conçu par les chercheurs du groupe IMP (Hakansson 1982 ; Valla, 1987), que nous nous proposons d'utiliser pour caractériser les interactions fournisseur/client, mais aussi les interactions entre fournisseurs, qui sont à l'œuvre dans le processus d'innovation qui nous intéresse. Cette caractérisation doit nous permettre de mieux cerner les rôles joués par les différents acteurs dans le processus d'innovation. La prise en

compte d'une multiplicité d'acteurs du processus d'innovation, tant chez le client, que chez les fournisseurs, peut participer de la mise en évidence d'une multiplicité des rôles joués notamment par les acteurs situés chez le client dans le processus d'innovation. Le caractère potentiellement multidimensionnel du rôle des premiers clients *lead users*, n'est pas pris en compte par le concept de *lead user*. Il se pose alors la question de savoir si le rôle de ces premiers clients dans le processus d'innovation étudié est réellement multidimensionnel, et si tel était le cas, si la prise en compte d'une telle caractéristique ne permettrait pas, alors, d'enrichir la notion de *lead user*.

## 2. L'INNOVATION ANALYSEE :

Nous avons effectué notre recherche en étudiant pendant plusieurs années le processus d'innovation conduisant à une innovation logicielle radicale multi-acteurs, de « création automatisée de base de connaissance ». Elle a été conçue et développée par deux start-up Temis et Mondeca, au travers de deux projets destinés aux clients PressPro et Exinis (les noms de ces deux sociétés ont été modifiés, pour des raisons de confidentialité). Temis a intégré, en 2004, une équipe issue du centre de recherche de Xerox, à l'origine d'une gamme de logiciels innovants dans le domaine du *text mining*, ou fouille de textes, dont le moteur d'extraction d'information IDE. Après cette acquisition, Temis a naturellement cherché à commercialiser ce logiciel en association avec des briques logicielles innovantes de son portefeuille de produits, elles-aussi de conception récente. Le modèle d'affaires de Temis devait initialement être celui d'un éditeur de logiciels packagés vendus « sur étagère », i.e. de progiciels, soit aux entreprises utilisatrices finales, soit aux sociétés de services informatiques chargées de leur intégration. Mais, le logiciel de Temis IDE, nécessitait la réalisation de développements personnalisés en vue de son adaptation à un domaine particulier (comme la veille technologique dans l'industrie pharmaceutique). Les sociétés de services informatiques et les premiers clients n'ayant pas les compétences pour réaliser de tels développements, Temis a dû assumer une activité de service, liée à leur réalisation.

C'est ensuite posé un autre problème, tout aussi classique dans l'industrie du logiciel. En effet, la majorité des premiers clients de Temis ne s'est pas contentée d'une solution de *text mining* fondée sur le logiciel IDE. Ces premiers clients recherchaient en général, une solution plus globale de *text processing*, ou « traitement automatisé de document », ce qui nécessitait d'associer d'autres logiciels à IDE, logiciels généralement développés par des start-up semblables à Temis. Il pouvait s'agir de moteurs de recherche professionnels, de systèmes de gestion de base de connaissance, de *crawlers* (robots d'indexation), ou de solutions de gestion de contenu. La connexion de différents logiciels

destinés à un même domaine d'application, pour réaliser une solution intégrée est généralement envisageable, sous réserve du respect de certains standards, mais exige dans la plupart des cas la réalisation de logiciels de type passerelle. C'est ainsi, qu'une application nouvelle couplant le moteur d'extraction d'information, IDE, de Temis, et le système de gestion de base de connaissance ITM, de Mondeca, a été conçue. Nous avons donc nommé cette application, solution de « création automatisée de base de connaissance ». C'est donc le processus d'innovation relatif à cette application, une innovation logicielle radicale multi-acteurs, que nous avons cherché à étudier.

Si l'étude de ce processus paraît de nature à répondre aux questions de recherche que nous avons fait émerger de la littérature, il faut souligner que le choix de ces questions a été lié à l'opportunité d'étudier un tel processus.

### **3. METHODOLOGIE :**

Nous avons tenté de répondre aux différentes questions que nous venons de préciser, en réalisant une analyse processuelle, conduite sur plusieurs années, portant sur une situation unique, que nous jugeons emblématique de l'industrie logicielle où l'innovation est généralement portée par des start-up technologiques, focalisées sur une brique logicielle spécifique. Les principales sources de données ont été constituées par les entretiens, que nous avons pu mener auprès des différents acteurs impliqués dans le processus d'innovation et par les documents que ces acteurs nous ont fournis.

### **4. PLAN DE THESE :**

Notre thèse se déroule selon le plan suivant :

- Au premier chapitre, tout d'abord, nous analysons les spécificités de l'industrie logicielle, marquée comme nous l'avons dit par le caractère immatériel et la modularité des architectures produit, mais aussi par un flux continu d'innovations radicales portées généralement par des start-up.
- Ensuite, au chapitre 2, nous étudions la littérature relative aux trois champs évoqués, étude d'où émergent des questions qui précisent notre question de recherche.

- Nous décrivons au chapitre 3, la méthodologie que nous avons retenue, et expliquons en quoi celle-ci paraît adaptée à notre question de recherche.
- Le chapitre 4 nous conduit, tout d'abord, à retracer le processus d'innovation étudié sur plusieurs années, au travers de la monographie d'un premier projet, le projet PressPro. Ce projet a permis de concevoir une architecture logicielle nouvelle, incluant une passerelle reliant le logiciel d'extraction terminologique de Temis à la base de connaissance de Mondeca. Nous décrivons cette passerelle, ainsi que les éléments permettant sa personnalisation. Puis, nous décrivons la méthodologie qui a été conçue pour organiser sa mise en œuvre. Nous présentons enfin une deuxième monographie, celle du projet Exinis, qui a vu l'implémentation complète de l'innovation et de la méthodologie qui accompagne sa mise en œuvre.
- Le chapitre 5 consiste en la discussion relative à ce processus d'innovation, et vise à répondre à la question de recherche, et aux différentes questions qui ont émergé de l'étude de la littérature. Cette discussion nous permet de caractériser un processus d'innovation et un processus de modularisation originaux. Cela nous amène à définir des concepts nouveaux, dans le champ de la modularité, concernant l'architecture produit. Cela nous conduit aussi à souligner le rôle capital des premiers clients dans le processus, et à enrichir la notion de *lead user*, en proposant quatre concepts nouveaux, qui prennent en compte le caractère multidimensionnel de la participation de ces clients.
- Enfin, la conclusion de ce travail de recherche, en montre certaines limites, et insiste sur l'intérêt de sa poursuite. En effet, au moment où nous avons terminé notre recherche, de nouvelles perspectives s'ouvraient quant à l'évolution de l'offre, quant à l'émergence de plateformes et de standards nouveaux, et quant au rôle des clients dans les futurs processus d'innovation.

# **Chapitre 1 : L'industrie du logiciel**

## SOMMAIRE DU CHAPITRE 1 :

<b>Chapitre 1 : L'industrie du logiciel</b>	<b>12</b>
<i>Sommaire du Chapitre 1 :</i>	<i>13</i>
1. <i>Introduction :</i>	<i>14</i>
2. <i>Une industrie jeune :</i>	<i>15</i>
3. <i>Le logiciel entre produit et services :</i>	<i>18</i>
4. <i>Les difficultés de l'innovation logicielle :</i>	<i>21</i>
5. <i>La création régulière de nouvelles entreprises et de nouvelles applications suivie d'un phénomène de concentration :</i>	<i>22</i>
6. <i>L'architecture logicielle et l'importance des concepts modulaires :</i>	<i>23</i>
7. <i>Stratégie des éditeurs de logiciel et effet de plate-forme :</i>	<i>24</i>
8. <i>Le Web sémantique, et l'émergence des applications de text mining :</i>	<i>36</i>

## 1. INTRODUCTION :

Selon un rapport de l'OCDE (1998), le logiciel est « le secteur le plus important économiquement, qui est à l'économie fondée sur le savoir ce que les secteurs de l'acier et de l'automobile étaient à l'économie industrielle ». Depuis, la tendance de l'économie et de la société à se « logicialiser » n'a fait que se renforcer, le rôle des technologies de l'information et de la communication (TIC) s'affirmant chaque jour un peu plus, et le logiciel prenant une place croissante au sein du secteur informatique au détriment des services et surtout du matériel informatique. Différents facteurs expliquent cette évolution, dont le développement des biens et services informationnels (ordinateurs, appareils et services de télécommunication,...) et la diffusion de biens et services dotés d'une dimension informationnelle importante, et ce dans tous les secteurs (électronique grand public, automobile, aviation, automates programmables, robots industriels, banque, etc.), évolution qui bénéficie de la généralisation de la numérisation des biens intangibles tels que les livres, les journaux, les photos, les films et la musique.

Le secteur des TIC recouvre des segments différents, dont l'industrie du logiciel (« *software* »), qui recouvre des activités industrielles et des activités de services. Cette industrie joue un rôle majeur, car elle est au cœur à la fois de l'innovation produit, de l'innovation dans les processus de production, mais aussi dans celle des organisations, des loisirs et de la vie quotidienne. L'édition de logiciels est résolument innovante : les 100 plus grands éditeurs français (*Truffle 100*, 2005) ont ainsi investi 26 % de leur chiffre d'affaires en R&D en 2005, soit 970 M€. Ce chiffre est supérieur au budget R&D de grandes entreprises industrielles telles que Schneider, L'Oréal ou Air Liquide, alors que chacune de ces entreprises est d'une taille comparable à l'ensemble de ces cent entreprises.

Le secteur du logiciel peut tout d'abord être appréhendé en distinguant les différents types de production, qui impliquent logiquement des acteurs différents, même s'il existe des recouvrements, du fait de l'importance des services associés aux logiciels :

- l'édition de logiciels ou produits logiciels (ou « *packaged software* ») ;
- le développement de logiciels spécifiques par des intégrateurs pour des entreprises industrielles et de services (services informatiques ou « IT services ») ;

- le développement, au sein des entreprises industrielles et de services, de logiciels correspondant à leurs besoins propres (le plus souvent des logiciels embarqués ou « *embedded software* »).

Ensuite, on distingue parmi les progiciels, trois grandes catégories, catégories décrites ci-dessous par taille de marché décroissante :

- les logiciels d'application ou applications : logiciels destinés à des activités transsectorielles (ex : ERP), à des activités sectorielles verticales (ex : CAO électronique), aux consommateurs finaux et aux applications domestiques.
- Les logiciels d'infrastructure système : systèmes d'exploitation, autres logiciels système, logiciels médiateurs ou *middleware*, logiciels de gestion de système et logiciels de sécurité.
- Les outils de développement de logiciels : systèmes de gestion de base de données ; composants, objets et environnements de développement ; gestion du cycle de vie de développement ; outils pour Internet.

En France, les éditeurs de logiciels étaient, en 2006, entre 1700 et 3400, avec plus de 30 000 salariés (selon rapport sur « l'Economie de l'Immatériel », coordonné par Lévy et Jouillet, 2006). La France possède de nombreux éditeurs de taille petite ou moyenne, et quelques leaders (Dassault Systèmes, Business Objects racheté par SAP en 2007, Ilog, Cegid,...). Mais globalement, le marché demeure très largement dominé par les entreprises américaines, et seule une entreprise française se place dans les vingt premières mondiales du secteur (Dassault Systèmes). L'environnement est particulièrement difficile car la compétition des grands acteurs mondiaux (SAP, Oracle, IBM,...) sur des segments jusqu'ici épargnés s'intensifie, tandis que la concurrence des « nouveaux » pays (Inde, Chine, Europe de l'Est...) s'affirme et que le secteur est entré dans un mouvement général de consolidation.

## **2. UNE INDUSTRIE JEUNE :**

L'industrie du logiciel est une industrie jeune et en évolution rapide, ce qui tend à compliquer son analyse. A l'origine, en effet, l'essentiel du logiciel était écrit par les utilisateurs eux-mêmes, car les constructeurs d'ordinateur n'en fournissaient pratiquement pas. Ainsi le premier ordinateur commercialisé par IBM, le 701, ne fut livré qu'avec une aide rudimentaire à la programmation, un mode d'emploi de cent pages, comprenant un assembleur primitif et quelques programmes de

gestion interne de la machine tenant sur une carte perforée (Armer, 1980). Les premiers utilisateurs devaient donc développer leurs propres logiciels. Cela conduisit à la création d'associations d'utilisateurs telles que la Digital Computer Association, puis le groupe SHARE, en 1955, dont le but était de partager l'information et les programmes entre utilisateurs de l'ordinateur IBM 704. La fourniture d'outils logiciels et de programmes d'assemblage par les constructeurs et les groupes d'utilisateurs ont permis de réduire le problème du coût de la programmation, mais la programmation d'application restait fastidieuse. Cela changea à l'apparition des langages de programmation du milieu des années 1950, comme le FORTRAN en 1957, devenu un standard de la programmation scientifique, tout comme le COBOL le deviendra plus tard, en informatique de gestion, et jusqu'aux années 1970 (Campbell-Kelly, 2003).

A la fin des années 1950, les constructeurs fournissaient les programmes nécessaires au bon fonctionnement de leur matériel, mais les programmes d'application étaient développés par les utilisateurs. Toutefois, progressivement, les entreprises furent tentées de confier l'écriture de leurs programmes d'application à des tiers, par manque de compétences en interne, ou par refus d'en développer. Ainsi le terme « *software industry* » apparut aux Etats-Unis, dans les années 1960, désignant le secteur proposant les éléments immatériels de l'informatique, logiciels et services. Depuis la fin des années 1960, ce terme désigne les entreprises commerciales qui produisent des logiciels.

L'industrie du logiciel est donc très récente, les produits logiciels ou progiciels destinés aux entreprises, principalement, n'étant apparu qu'au milieu des années 1960, après le lancement des ordinateurs IBM de la gamme 360, qui ouvrit un marché d'une taille importante pour l'époque, et surtout après que, sous la pression de la législation antitrust américaine, IBM décida de facturer séparément la partie « services » et la partie « logiciel ». Ce dégroupage donna un coup de fouet au marché des progiciels, qui était balbutiant à cette époque. L'explosion du marché des micro-ordinateurs à partir des années 1980, permit pour sa part le développement de logiciels édités pour un très grand nombre de clients (entreprises ou individus), et distribués sous une forme « packagée » (progiciels) dans des magasins ou par correspondance.

Comme nous l'avons vu, le logiciel voit son importance grandir (Stryszowski, 2008) en ce qu'il est un composant important d'un nombre croissant de produits ou de processus, pour les entreprises comme pour les individus. L'innovation dans le secteur logiciel se caractérise d'une part, par le fait qu'il s'agit d'un secteur hautement innovant en tant que tel, et d'autre part par le fait qu'il

constitue un élément important d'innovation dans d'autres secteurs. L'innovation dans l'industrie logicielle se caractérise notamment par des cycles de vie courts, et un effet réseau marqué (sur lequel nous reviendrons dans les points suivants). Le processus d'innovation dans le logiciel se caractérise quant à lui par un besoin relativement faible en capital, des coûts marginaux de production négligeables (importance du piratage, et de la position dominante), l'innovation sur une base cumulative, et la spécificité des modes de valorisation de l'innovation (copyright, logiciels open source). Les incitations à adopter une innovation logicielle sont principalement l'augmentation de la productivité, l'amélioration du service fourni, ou encore la réduction des coûts. Un environnement de R&D favorable à l'innovation repose sur la disponibilité du capital humain, un stock de connaissance publique (ex. des recherches conduites dans le secteur public, ou reposant sur des liens forts entre la recherche et les entreprises), le support des marchés financiers (dont la disponibilité d'un support financier public direct pour la R&D privée). Par ailleurs, les barrières à l'entrée sont relativement basses, elles sont constituées des menaces potentielles sur les prix, des restrictions d'accès aux canaux de distribution, des accords de vente exclusive et des barrières légales. Par contre, la durée de vie d'un logiciel peut être longue, elle passe par la coopération avec des développeurs de matériel et d'autres technologies, et par une bonne protection des droits de propriété intellectuelle (pour les logiciels commerciaux, elle assure un flux régulier de profit, pour les logiciels open source, elle empêche l'expropriation).

Horn (2004) observe que, les producteurs de logiciel peuvent avoir des activités principales très diverses. A côté des éditeurs de logiciel focalisé sur cette activité, on trouve en effet des sociétés de services, des constructeurs de matériel informatique, des opérateurs de télécommunication, et des entreprises industrielles et de services qui intègrent des logiciels dans leurs produits. Cela résulte de la désintégration de l'industrie informatique, ainsi que de sa spécialisation. Depuis les années 1980, la place des entreprises généralistes, capables de fournir matériels, logiciels et services informatiques a en effet diminué au profit de spécialistes. Cette situation peut s'expliquer tout d'abord par l'émergence de la mini-informatique, puis de la micro-informatique, univers dans lequel des entreprises spécialisées se sont imposées. La spécialisation a un stade précis de la filière informatique (composants, assemblage de matériel, périphériques, logiciels, services,...) a débuté avec la création des SSII. La désintégration verticale s'est accélérée avec le développement de la microinformatique, constituée immédiatement de spécialistes fournissant chacun un élément du système, à l'exception notable de l'univers Apple. L'existence de telles entreprises spécialisées, est le résultat logique de conditions technologiques et de marché très différents aux différents niveaux de la filière : compétences requises, durée de cycle de vie des produits, exigences

organisationnelles, dynamique concurrentielle. Les entreprises nouvellement créées se focalisent souvent sur une activité spécifique, en cherchant à valoriser parfois une avancée technologique particulière. Dans le cas du logiciel destiné aux entreprises, la focalisation porte même sur un type de produit en particulier, tendance qui n'est que faiblement contrecarrée par la recherche de solutions globales par les entreprises clientes. En effet, l'activité d'intégration de systèmes a été prise en charge par des sociétés de services, les SSII. Dans le secteur des progiciels, la majorité des entreprises se spécialisent sur un type de matériel et/ou sur une activité précise. On assiste aussi, depuis plus de dix ans, à la volonté des grands constructeurs informatiques, comme IBM, de développer leur activité logicielle, de laquelle ils s'étaient désengagés dans les années 1980. Par ailleurs, les perspectives de profit très élevées incitent de grandes entreprises à racheter des entreprises de logiciel. De plus, des éditeurs de logiciel installés tentent de se diversifier, souvent par acquisition externe dans l'informatique. Mais, en synthèse, la tendance dominante reste clairement à la spécialisation et à la désintégration de l'industrie informatique.

### **3. LE LOGICIEL ENTRE PRODUIT ET SERVICES :**

Le logiciel est un produit susceptible d'occuper une position variable sur le continuum des biens et des services (Horn, 2004). En effet, les logiciels peuvent être considérés comme des biens intangibles ou des services. Ils peuvent aussi être encadrés dans des biens tangibles, puisque dans certains cas ils sont intégrés de manière indissociable aux composants électroniques qu'ils pilotent. Mais, en général, un logiciel existe indépendamment des composants sur lequel il agit, et l'on distingue, comme nous l'avons déjà évoqué :

- le progiciel, logiciel développé sous une forme standard par un éditeur de logiciel, qui entend satisfaire aux besoins de nombreux utilisateurs : « ensemble complet et documenté de programmes conçu pour être fourni à plusieurs utilisateurs en vue d'une même application ou d'une même fonction » (Journal Officiel). La proportion des progiciels par rapport à l'ensemble des logiciels, biens intangibles, croît régulièrement, passant de 5,5% en 1992 à 9% en 2002 (OCDE, 2002).
- le logiciel sur-mesure développé en réponse à un besoin spécifique, par le service informatique de l'entreprise qui en aura l'usage ou par une société de services informatiques.

Certains auteurs (Campbell-Kelly, 2003) font une distinction entre les progiciels destinés aux entreprises exclusivement, comme les ERP (Enterprise Resource Planning) ou, en français, PGI (Progiciel de Gestion Intégrée), dont l'implantation nécessite souvent l'intervention d'une SSII, et les progiciels destinés au marché de masse, comme la suite Microsoft Office. Dans le premier cas, le modèle économique adopté par les développeurs de tels logiciels est celui de producteur de biens d'équipement (justement comme les ordinateurs), dans le deuxième cas, c'est le modèle de producteur de biens d'information, celui de l'industrie du disque par exemple.

Les années 1960 virent le lancement de milliers d'entreprises proposant des services de programmation, aux Etats-Unis, en Europe et au Japon. Mais la production de logiciel sous contrat étant une activité cyclique, les grandes SSII se diversifièrent dans les services de traitement et l'infogérance. La production de logiciel sous contrat peut être caractérisée par de faibles barrières à l'entrée, et par un taux de mortalité élevé, du fait de la concurrence sur les prix et de la nature cyclique de la demande.

Comme nous l'avons déjà évoqué, l'émergence des produits logiciels date de la deuxième moitié des années 1960, et résulte principalement de quatre facteurs : la prolifération des ordinateurs, l'évolution du rapport des coûts entre logiciel et matériel, la « crise du software » (manque de développeurs informatiques) et enfin l'introduction d'une plate-forme standard, l'IBM 360. La facturation séparée du matériel et du logiciel par IBM, au début des années 1970, sous la pression de la législation antitrust aux Etats-Unis, a aussi joué un rôle quant au développement de ce marché. Le concept de package logiciel évolua : d'abord, le logiciel « prépackagé » fut une réponse des constructeurs à la croissance exponentielle d'ordinateurs vendus, alors que le nombre de programmeurs évoluait linéairement. Le package logiciel permettait de faire sauter ce goulet d'étranglement. Mais les packages gratuits fournis par les constructeurs ne suffisant plus, un marché des packages logiciels apparut vers 1965, et les SSII introduisirent le produit logiciel ou progiciel pour exploiter cette opportunité. Les produits logiciels destinés aux entreprises, se sont avérés plus coûteux que ne l'avaient espéré ces SSII : du fait de leurs coûts commerciaux élevés, des charges d'avant et d'après-vente, de la nécessaire maintenance et du développement de nouvelles versions, et finalement ces progiciels sont apparus peu différents des biens d'équipement classiques.

Le secteur du logiciel commercialisé auprès de moyennes et grandes entreprises aura probablement toujours une composante « service » importante (Cusumano, 2004). La croissance

du marché des progiciels n'a pas contribué à réduire la demande de services liés au logiciel. Les besoins des entreprises clientes sont rarement complètement comblés par les logiciels packagés, ce qui crée un marché énorme et régulier, pour les logiciels personnalisés et les services liés à l'intégration informatique. Pour Cusumano (2004), la décision fondamentale qui incombe aussi bien aux start-up qu'aux firmes bien établies dans le champ du logiciel d'entreprise, ne se limite pas à choisir entre être une société de services ou un éditeur de progiciel, mais consiste à déterminer la place que son activité de services doit occuper dans son activité globale. De plus, à différents moments de l'évolution de ses lignes de produit et de sa base client, une entreprise de logiciel fera évoluer l'accent qu'elle mettra sur l'une ou l'autre de ces activités. Les modèles d'affaires peuvent aussi évoluer en fonction de la conjoncture. Plus l'entreprise et ses produits sont matures, plus elle est susceptible de mettre l'accent sur son activité de services pour générer des revenus récurrents. Quand la conjoncture freine les investissements technologiques, ou quand les anciens produits ne se vendent plus très bien, les services peuvent s'avérer une source de revenu indispensable.

Ensuite, il faut souligner que des changements dans les technologies « plate-forme » (cf. le rôle des « *platform leaders* », que nous abordons plus loin dans ce chapitre) génèrent des demandes nouvelles, relatives à de nouveaux produits et à de nouveaux services permettant aux clients de réutiliser leurs actifs logiciels. Dans le passé, les logiciels système destinés aux nouvelles plates-formes, spécialement les systèmes d'exploitation, représentaient des marchés horizontaux énormes et des flux de revenus récurrents, pour les firmes qui dominaient ces segments. Mais les possibilités offertes aux nouveaux entrants étaient très limitées : les constructeurs d'ordinateurs et quelques éditeurs de logiciel établissent en général le standard de la plate-forme et sont extrêmement difficiles à déloger. Néanmoins les fournisseurs de hardware sont rarement capables de couvrir tous les segments de logiciel, et parfois n'ont pas la capacité à développer tous les logiciels systèmes pour une nouvelle plate-forme. Il arrive aussi que ces firmes ne sachent pas traiter le logiciel comme un secteur commercial à part entière. Cela les rend vulnérables à la concurrence d'entreprises dédiées à la production de logiciel, sur des niches de marché particulières.

Enfin, l'histoire du marché du logiciel, montre que les niches d'applications et les nouvelles plates-formes constituent les meilleurs endroits à explorer pour trouver de nouvelles opportunités de services et de produits logiciels. Au début des années 2000, plus que les logiciels système destinés au matériel standardisé, les niches d'application horizontales et verticales (bases de

données, ERP, etc.) et les nouvelles plates-formes informatiques telles que téléphones portables connectés à Internet, les ordinateurs portables et les Web services, ont ainsi fourni des opportunités remarquables.

#### **4. LES DIFFICULTES DE L'INNOVATION LOGICIELLE :**

Malgré sa jeunesse, le secteur des logiciels se caractérise par un rythme élevé d'apparition d'innovations, de tous ordres. Ce secteur a en effet connu des innovations majeures. De nouvelles applications sont apparues et se sont banalisées rapidement, comme le courrier électronique, par exemple, tandis que des innovations fondamentales mais moins visibles aux yeux des utilisateurs ont révolutionné le secteur, comme dans le domaine de l'architecture des systèmes informatiques, avec le passage d'une architecture « maître-esclave » reposant sur des *mainframes* à une architecture « client-serveur », de type distribuée. A côté des innovations radicales, le secteur se caractérise par un flux continu d'innovations incrémentales. Cela résulte de ce que le logiciel est aisément modifiable, ce qui permet théoriquement de le personnaliser, de l'adapter aux besoins particuliers de certains utilisateurs, mais aussi de ce que l'environnement de tout logiciel installé évolue rapidement, que ce soit au niveau des matériels sur lesquels il fonctionne ou que ce soit au niveau des autres logiciels avec lesquels il doit s'interfacer (par exemple, les systèmes d'exploitation), ce qui nécessite des adaptations et mises à jour régulières.

Mais si la production de logiciels est une activité fortement innovante, elle suscite de manière récurrente le mécontentement des utilisateurs (Cusumano, 2004 ; Horn, 2004). Une constante dans l'histoire de l'industrie du logiciel, est en effet la difficulté d'écrire et de tester de nouveaux systèmes logiciels, spécialement pour de nouvelles plates-formes et de nouvelles applications. Comment créer un processus de développement logiciel qui s'accommode des changements technologiques rapides, et des demandes incertaines des clients ? Des tentatives pour résoudre les problèmes récurrents rencontrés pour développer des logiciels innovants ont été faites pour structurer la technologie au travers de « software factories », une approche « industrielle » comme son nom l'indique, et au travers d'efforts dans le champ de l'ingénierie logicielle. L'approche « software factory » appliquée principalement au Japon et aux USA a été largement critiquée, même si certains y ont vu un progrès par rapport à l'artisanat pratiqué auparavant. De toute façon cette approche a été marginalisée par d'autres, ainsi la philosophie Microsoft incite à alterner de phases de synchronisation, et des phases de stabilisation. La combinaison de techniques anciennes et récentes, a permis de bâtir des progiciels de plus en plus complexes, et de répondre

rapidement à des changements technologiques et de marché. Mais l'idée première des tenants de l'approche « *software factory* », selon laquelle aucun projet n'est entièrement nouveau, et qu'il ne faut pas repartir d'une feuille blanche à chaque fois, reste vraie. Pour Cusumano (2004), le logiciel n'est fondamentalement pas une activité manufacturière, mais plutôt une forme de conception de produit, où le concept est le produit et la réplication triviale.

En synthèse, les problèmes de dépassement de délais des processus de conception et de développement et l'insatisfaction récurrente des utilisateurs finaux ont conduit à l'adoption de méthodes itératives. De telles méthodes encouragent les interactions entre concepteurs et utilisateurs tout au long du processus, et plus seulement lors des phases amont de définition des besoins et des spécifications du logiciel et aval de validation du logiciel ; ces méthodes permettent aussi le retour aux étapes précédentes du processus d'innovation, si nécessaire.

## **5. LA CREATION REGULIERE DE NOUVELLES ENTREPRISES ET DE NOUVELLES APPLICATIONS SUIVIE D'UN PHENOMENE DE CONCENTRATION :**

Les éditeurs de logiciel sont divers à plus d'un titre, et notamment par leur taille (Horn, 2004). D'un côté le marché est très concentré puisque les vingt plus grands fournisseurs mondiaux de logiciels et de services connexes représentent plus du tiers du marché mondial, et ce depuis 2000. Ce phénomène de concentration s'est même accéléré depuis une dizaine d'années. De plus, pris séparément, le secteur du progiciel apparaît comme encore plus concentré : les dix plus grands éditeurs de logiciel réalisent plus du tiers des ventes de logiciel dans le monde. Ensuite, il faut tenir compte du fait que le marché du progiciel est éclaté en de nombreux sous-segments, où l'on observe une situation oligopolistique, voire parfois monopolistique. La principale raison tient aux rendements croissants d'adoption, dont font partie les effets de réseau sur lesquels nous revenons au point 7 de ce chapitre. Il faut souligner que ces rendements sont particulièrement importants pour les progiciels : en effet, l'apprentissage par l'usage est un élément essentiel de l'assimilation des logiciels, et les interrelations technologiques sont nombreuses entre matériels et logiciels, et entre logiciels de différents types. Les économies d'échelle sont bien sûr évidentes, et les rendements croissants d'information sont importants pour les progiciels dont il est toujours difficile de définir a priori les rendements réels, ce qui entraîne des phénomènes d'imitation et de réputation. Enfin la diffusion d'un logiciel s'appuie sur des externalités de réseau fortes. Ces phénomènes conjugués à une baisse régulière des prix, font qu'un progiciel pourra s'imposer relativement rapidement sur un segment particulier, voire parfois dominer seul le marché.

Ainsi un phénomène de concentration du marché apparaît, tandis que d'autres petites entreprises émergent sur de nouveaux segments (nouvelles applications ou nouveaux types de matériel par exemple), où les barrières à l'entrée sont faibles. Cela explique le très grand nombre de petits éditeurs de logiciel, les trois quarts ayant même moins de dix employés. Ainsi en France, il s'est créé 4000 entreprises par an dans ce secteur pendant les années 1990, alors que le nombre total restait stable autour de 20000 (Insee). Certaines de ces start-up réussissent rapidement, ce qui est illustré par le fait que 40% des 500 premières entreprises en 1999 avaient moins de dix ans (OCDE, 2002). Ceci explique la part importante du capital-risque qui est captée par les éditeurs de logiciel. Ainsi, dans l'économie du logiciel voit-on coexister de très grandes entreprises, généralement en position dominante sur un ou plusieurs segments de marché, et un grand nombre de petites entreprises spécialisées sur des nouvelles applications. La consolidation ultérieure étant compensée par l'éclosion récurrente de nouveaux segments (Horn, 2004).

## **6. L'ARCHITECTURE LOGICIELLE ET L'IMPORTANCE DES CONCEPTS MODULAIRES :**

Un « module » dans le cas du logiciel, peut être défini comme un sous-ensemble de fonctionnalités, qui est plus petit que le produit entier, et que les concepteurs peuvent isoler d'autres petits ensembles de fonctionnalités, et dans une certaine mesure, tester comme unité indépendante. Une architecture modulaire, à la différence d'une architecture intégrée permet de modifier et de tester des parties d'un produit sans créer de bogues dans les autres composants. Certains logiciels de petite taille peuvent avoir des architectures intégrées, pour minimiser justement leur taille et optimiser leurs performances. Mais au fur et à mesure que la taille du code grossit, jusqu'à atteindre des millions de lignes de code, il devient indispensable d'adopter une architecture modulaire.

Une architecture modulaire définit ce que doivent être les sous-systèmes (constitués de différents modules) du produit, comment les sous-systèmes sont connectés entre eux, ce qu'est un module à l'intérieur d'un sous-système, et surtout ce que doivent être les interfaces pour que les sous-systèmes et les modules individuels puissent travailler ensemble. Les interfaces doivent rester stables pendant un certain laps de temps, parce que les développeurs doivent absolument savoir comment faire pour que les modules interagissent entre eux. La modularisation aide aussi une équipe-projet à prioriser les fonctionnalités, en fonction de leur importance en termes techniques et de marché. La priorisation et la modularisation donnent à l'équipe la possibilité de supprimer

certaines fonctionnalités, en cas de retard. A l'inverse, l'adoption d'un processus séquentiel, conduit le projet à suivre une planification de type «*waterfall*», et ne permet pas de tester les différentes parties d'une manière intégrée, tant que le travail n'est pas complètement terminé, ce qui empêche de résoudre les problèmes majeurs, ou de faire des changements importants pour le client. C'est pourquoi ce type d'approche n'est généralement pas retenu, et les architectures modulaires se sont imposées dans le logiciel, depuis les premiers travaux de Parnas sur le sujet (1972). On observe ainsi que certains éditeurs de logiciel ont fait évoluer des architectures produit vers plus de modularité, comme ce fut le cas de Microsoft avec l'ensemble des logiciels qui constituent la suite Office : Word, Excel, Powerpoint et Access. Au départ, ces produits étaient totalement indépendants dans leur conception, et une équipe de développement complètement indépendante était en charge de les relier horizontalement. Par la suite, Office est devenu le produit, et Word, Excel et Powerpoint sont devenus des sous-systèmes partageant la moitié de leur code. De même, Windows a été partiellement reconçu et a évolué vers une architecture modulaire, ce qui n'était pas le cas initialement (Cusumano, 2004).

## **7. STRATEGIE DES EDITEURS DE LOGICIEL ET EFFET DE PLATE-FORME :**

Dans l'industrie, on observe souvent que le leader d'un segment de marché arrive à acquérir une position très dominante, une situation où le « vainqueur prend tout » (« *winner takes all* »). Cela résulte notamment de l'importance des standards (émergence de logiciels standards dans un domaine d'application donné qui permet le partage et l'échange de fichiers entre entreprises, par exemple : SAP dans les ERP, Microsoft dans les systèmes d'exploitation ou les suites bureautiques, Dassault Systèmes/ PTM/UGS dans les logiciels PLM, etc...), de l'effet de réseau (déjà évoqué) et de l'effet de plate-forme.

### **Quelle stratégie pour les acteurs d'une « *winner takes all industry* » ?**

Dans le rapport sur l' « Economie De l'Immatériel », Levy et Jouillet (2006) analysent les dynamiques de marché dans l'économie de l'immatériel. Selon eux, un des facteurs de risque et d'incertitude dans l'économie de l'immatériel est lié au processus d'innovation en tant que tel et à ses conséquences sur le fonctionnement des entreprises. En effet, dans un premier temps, l'innovation conduit à des « monopoles temporaires » et génère des effets très discriminants de réseau et de prime au premier entrant, ce qui, en soi, rend la concurrence particulièrement vive et la probabilité d'échec assez élevée. Mais dans un deuxième temps, le processus d'innovation

remet en cause ces rentes temporaires, remettant ainsi quasiment au même plan tous les différents acteurs. D'abord, l'économie de l'immatériel est marquée de plus en plus par des effets de réseau. Un effet de réseau peut être défini de la façon suivante : chaque individu bénéficie d'autant plus de sa participation à un réseau que le nombre de personnes qui y sont également connectées est important : plus un réseau est important, plus ses membres peuvent en tirer des bénéfices. Ces effets de réseau existent dans de nombreux secteurs : les logiciels (avec notamment les systèmes d'exploitation), les télécommunications, Internet (exemple des sites eBay ou YouTube) ou encore les standards électroniques. Les entreprises doivent donc prendre en compte ces effets de réseau qui peuvent à la fois réduire le risque, lorsque l'entreprise se trouve au cœur du réseau ou, au contraire, fortement l'accroître, lorsqu'il s'agit de concurrencer une entreprise qui en bénéficie. L'existence de ces effets de réseau induit donc une dynamique concurrentielle particulière, parfois qualifiée d'effet *winner-takes-all*. En effet, l'entreprise qui parvient à devenir le réseau de référence capte l'essentiel des gains alors que ses concurrents voient leurs investissements très largement perdus. L'exemple caractéristique de cet effet peut se trouver dans l'industrie du logiciel avec Microsoft et le système d'exploitation Windows. Des travaux de recherche ont d'ailleurs confirmé cet effet dans le secteur de l'Internet (Hand, 2001).

Ces effets de réseau sont des externalités de consommation qui proviennent de l'interdépendance de décisions de consommation (Horn, 2004). La valeur du bien augmente avec le nombre d'utilisateurs. Ces externalités de réseau peuvent être directes, comme dans le cas du téléphone (plus il y a d'abonnés, plus il y a d'intérêt à l'être), ou indirectes, quand, par exemple, l'offre de produits complémentaires dépend de la taille du réseau (ex : quand le nombre de jeux dépend du type de console choisie). Les externalités de réseau se distinguent des autres rendements croissants d'adoption, en ce que, dans le cas des externalités de réseau, les rendements associés à une technologie sont aussi associés au comportement des utilisateurs futurs, et pas seulement au comportement des utilisateurs passés de cette technologie. Ainsi, l'utilisateur va chercher à adopter la technologie qui l'emportera à terme, ce qui est toutefois généralement difficile à déterminer. Ceci peut, par exemple, l'amener à choisir un produit qui sera compatible avec un standard émergent, comme le sera peut-être l'architecture UIMA, promu par IBM, pour le *text mining*, comme nous le verrons plus loin (p. 39).

La combinaison d'effet de réseau du côté de la demande, et d'économies d'échelle du côté de l'offre peut conduire une entreprise à détenir un monopole (Arthur, 1996; Shapiro et Varian, 1999; Leibowitz et Margolis, 2000). C'est ce que Arthur (1989) a nommé les Rendements

Croissants d'Adoption (RCA), concept selon lequel une technologie n'est pas choisie parce qu'elle est la meilleure, mais devient la meilleure parce qu'elle est choisie. Son argumentation est la suivante : par l'effet conjugué des économies d'échelle (réduction du coût unitaire en fonction du volume de production), de l'apprentissage par la pratique (amélioration des performances par l'expérience) et des externalités (ou effets) de réseau, on parvient au résultat selon lequel plus un nouveau produit technologique est « adopté », plus ses coûts de production baissent et son utilité augmente, et cela de façon plus que proportionnelle. Les rendements d'adoption sont alors dits « croissants ». Arthur (1989) a ainsi mis en évidence des mécanismes d'auto-renforcement qui se manifestent autour d'une technologie. Les sources des rendements croissants d'adoption sont l'apprentissage par l'usage, les interrelations technologiques, les économies d'échelle en production, et les effets (ou externalités) de réseau.

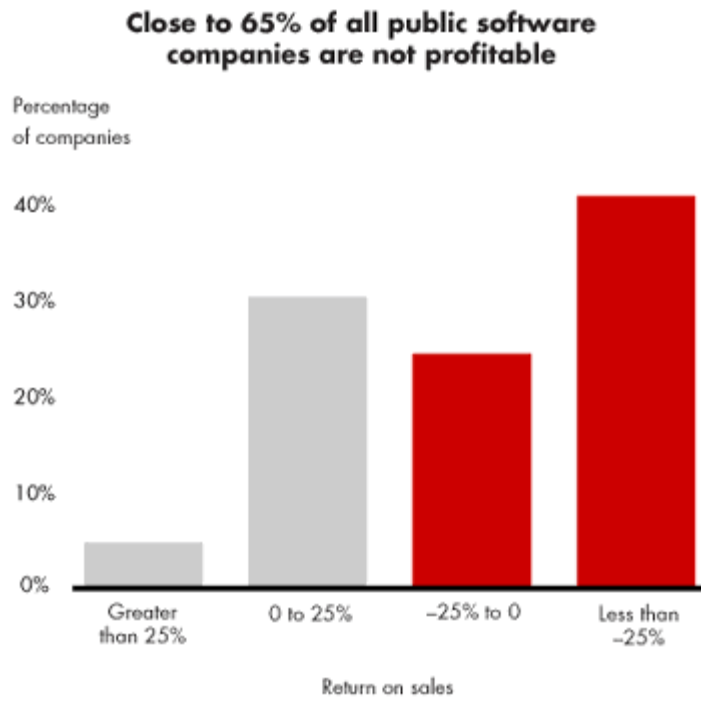
Une conséquence économique majeure est donc que l'efficacité économique à travers les mécanismes de marché n'est plus nécessairement assurée : une technologie « sous-optimale » peut s'imposer. Les logiciels, produits intangibles fonctionnant sur des ordinateurs, eux-mêmes en réseau, constituent un univers où économies d'échelle (côté offre) et effet de réseau (côté demande) jouent un rôle majeur. Grâce à cet effet, une base clients de grande taille a tendance à croître fortement parce que les clients actuels attirent de nouveaux clients, même en l'absence de dépenses engagées par l'entreprise (Mauboussin, Schay et Kawaja, 2000). Cet effet de réseau joue un rôle important dans l'univers du logiciel en général, qu'il soit destiné ou non à Internet. Par ailleurs, l'importance des économies d'échelle est évidente dans l'industrie du logiciel, où les coûts fixes d'investissement dans la production de logiciels sont élevés, alors que les coûts marginaux sont pratiquement nuls (Woodall, 2000; Noe et Parker, 2000).

La conscience de ces effets (effet de réseau et économies d'échelle), ainsi que les parts de marché et l'étonnante rentabilité d'éditeurs de logiciel comme Microsoft ou SAP, ont conduit la plupart des éditeurs de logiciels innovants à adopter une stratégie « *winner takes all* ». En conséquence, ces entreprises ont tendance à dépenser énormément d'argent pour embaucher du personnel, développer des logiciels innovants, voire promouvoir leur marque, afin de prendre une longueur d'avance et d'acquérir une position dominante sur leur marché. En effet, un autre élément vient compléter et renforcer les effets de réseau et d'économies d'échelle, c'est la prime au premier entrant ou *first-mover advantage* : la première entreprise qui propose le produit ou le service concerné peut obtenir un avantage concurrentiel très important et difficile à remettre en cause par ses concurrents arrivés plus tard sur le marché. Les effets de réseau sont une des explications

du *first-mover advantage* mais celui-ci est plus large : dans le cadre d'une innovation ou d'un nouveau marché, l'entreprise qui propose en premier le nouveau produit ou le nouveau service attire spontanément les consommateurs qui supposent qu'elle est la mieux à même de fournir ce produit ou ce service (naturellement, mis à part les effets de réseau proprement dits, la prime au premier entrant est généralement transitoire et peut même, dans certains cas, ne pas exister).

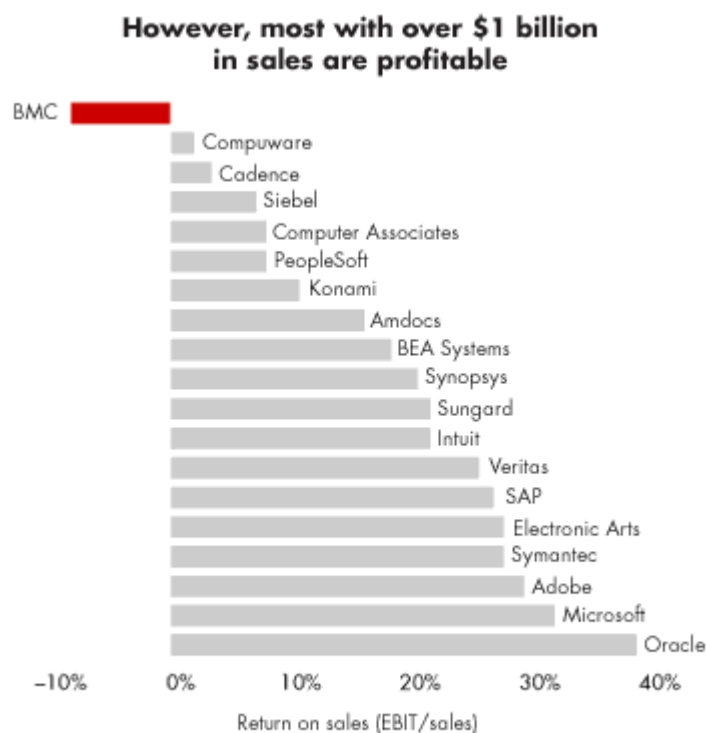
Comme nous venons de le voir, la combinaison de ces différents effets conduit à des processus d'innovation et de mise sur le marché assez risqués, dans lesquels les entreprises doivent d'abord faire face à des coûts fixes extrêmement élevés (R&D, publicité...) pour rapidement entrer sur le marché, avec le risque qu'un concurrent plus rapide occupe déjà une position très forte voire exclusive, et sans possibilité, en cas d'échec, de récupérer au moins une partie de ces investissements. C'est d'ailleurs ce phénomène qui a été observé à la fin des années 1990 au moment de la « bulle Internet ». Face à l'explosion du commerce électronique, de nombreuses start-up se sont efforcées d'acquérir une prime au premier entrant en développant leur site Internet et en investissant massivement dans la publicité pour se faire connaître et créer une marque de référence. Un des meilleurs exemples réussis de cette stratégie est certainement le site de vente en ligne Amazon qui, en arrivant parmi les tous premiers sur la vente de livres en ligne, est devenu une marque de référence puis a largement élargi le champ des produits vendus. En revanche, l'échec de très nombreuses start-up et l'effondrement de la bulle Internet à partir de 2000 ont montré qu'en dépit de l'importance des investissements réalisés, la « nouvelle économie » de l'immatériel restait marquée par un niveau d'incertitude très élevé.

Ainsi, une analyse réalisée en 2004 par Bain & Company, montre qu'il n'y a pratiquement que deux sortes d'entreprises de logiciel qui font des bénéfices : les leaders d'une catégorie, et ceux dont le chiffre d'affaires dépasse le milliard de dollars (Figure 1, et Figure 2 page suivante). En effet, comme on peut l'observer sur le diagramme ci-après, près des deux tiers de ces entreprises n'avaient pas fait de profit en 2003. Dans le même temps, et grâce à la croissance de leur titre en bourse, les grandes entreprises ont amassé des trésors de guerre, tout en se rendant compte que la vente de nouvelles licences logicielles à leurs clients ne permettait pas une croissance significative. Cela les conduit à envisager des rachats d'éditeurs de logiciel, présents sur d'autres marchés que les leurs.



Source: Compustat, Bloomberg, CY2003 data; Bain analysis.

Figure 1 : Rentabilité des éditeurs de logiciel. (Bain, 2004).



Source: Compustat, Bloomberg, CY2003 data; Bain analysis.

**Figure 2 : Rentabilité des éditeurs de logiciel dépassant le milliard de dollars de chiffre d'affaires annuel (Bain, 2004).**

En 2004, les analystes de Bain estimaient qu'une évolution fondamentale du marché allait conduire à hâter la concentration de l'industrie logicielle. Les dépenses des entreprises dans le domaine des technologies de l'information avaient en effet baissé, après l'éclatement de la bulle Internet. La tendance du marché était alors à aller vers un environnement informatique plus fédéré, alors que la situation antérieure était plutôt caractérisée par des produits faiblement intégrés, et à aller vers des architectures informatiques conçues autour de standards ouverts ou mises en place par des consortiums de grands éditeurs de logiciel. Les clients commencèrent à moins rechercher les solutions hyper-performantes, qu'elles doivent faire coexister à leurs propres frais, et à s'orienter vers plus de fiabilité et de simplicité. Cette étude prévoyait donc une consolidation globale de l'industrie logicielle, tout en soulignant que les fusions-acquisitions dans le secteur des entreprises technologiques ont souvent détruit de la valeur, et finalement échoué. Cette étude concluait d'une part, que dans l'industrie du logiciel, les plus petits acteurs sont presque tous, à un moment ou à un autre, confrontés à des choix très délicats : fusionner avec un autre acteur (en étant parfois absorbé par un gros éditeur), essayer de devenir un logiciel plateforme, notion sur laquelle nous revenons au point suivant, créer ou se maintenir sur des niches de marché, souvent en tant qu'acteur de complément, notion que nous reprenons aussi au point

suivant. Certains peuvent aussi envisager de se développer dans les services connexes, alors que leur chiffre d'affaires provenant des ventes de licences logicielles décline. D'autre part, la conclusion de l'analyse était que les grandes firmes, qui réalisent un chiffre d'affaires supérieur au milliard de dollars, peuvent, quant à elles, faire des acquisitions en étendant leur champ d'action à des catégories adjacentes, et en captant de grandes bases installées. Parmi d'autres méthodes, ces entreprises peuvent aussi envisager de travailler avec de petits acteurs, comme revendeurs ou partenaires OEM (i.e. dont une brique logicielle est intégrée de manière transparente à l'offre logicielle de la grande entreprise) avant de les absorber. Ils peuvent aussi se développer par la verticalisation de leur offre, comme par exemple Siebel avec ses packages logiciels destinés au management des relations clients dans l'industrie pharmaceutique. La question des fusions-acquisitions dans l'univers du logiciel, est aussi tributaire de la régulation sur l'abus de position dominante, à laquelle Microsoft, par exemple, est régulièrement confronté. Cela nous conduit à explorer la notion de logiciel plate-forme, et son importance dans l'industrie du logiciel.

#### « *Platform leadership* » :

Comme nous l'avons évoqué au point précédent, dans l'industrie du logiciel, les éditeurs qui définissent leur stratégie ont en fait à choisir en trois orientations principales : être leader (catégorie dans laquelle on trouve les « platform leaders »), être suiveur, ou être acteur de complément (« complementor », Gawer et Cusumano, 2002 ; Cusumano, 2004).

La notion de platform leader a été illustrée par Andy Grove (1996), CEO d'Intel, et développée par Gawer et Cusumano (2002). Un *platform leader* est une entreprise dont les produits servent de fondation à d'autres entreprises (« *complementors* ») pour construire leurs propres produits ou proposer leurs propres services. Les *platform leaders*, ou leaders de plate-forme, sont donc les acteurs qui peuvent établir les standards d'un marché, qu'il s'agisse d'un marché de masse ou d'une niche. Sur certains segments de marché, comme celui des systèmes d'exploitation, être leader implique d'être *platform leader*. Les produits complémentaires, fournis par les *complementors*, comme les périphériques pour micro-ordinateur ou les logiciels d'application sont souvent nécessaires pour que les plates-formes logicielles, comme Windows (ou matériel, comme les microprocesseurs Intel) rencontrent le succès. Pour Grove, les *complementors* constituent même une sixième « force de Porter ». Les leaders de plate-forme peuvent souvent tirer leur épingle du jeu même dans le cas d'un marché en récession, car les clients continuent généralement à acheter des nouveaux « produits de plate-forme », alors même qu'ils réduisent leurs autres dépenses. Les

entreprises « complémentaires », comme Dell dans l'univers du PC, mettent tous leurs efforts pour vendre leurs produits, ce qui conduit mécaniquement à augmenter les ventes de produits plates-formes, comme les micro-processeurs Intel et le système d'exploitation Windows de Microsoft. Cusumano observe qu'en général, une entreprise de logiciel choisit au départ soit de devenir un leader technologique, soit de devenir un leader de plate-forme, mais la majorité finit par devenir un suiveur et un acteur de complément de la plate-forme d'un autre acteur du marché.

En effet, les leaders technologiques ne sont pas forcément des leaders du marché. Des exemples célèbres viennent illustrer ce propos, comme celui de Xerox, qui participa aux premiers développements relatifs au photocopieur, dans les années 1950, mais fut ensuite dépassé par les fabricants japonais dans les années 1980, qui introduisirent des produits bon marché mais très fiables. En informatique, il y a de nombreux exemples de firmes qui échouèrent à tirer avantage de leur R&D et de leurs innovations, et ne parvinrent jamais à devenir les leaders de leur marché. Là encore, on peut citer Xerox qui est, au moins en partie, à l'origine d'innovations telles que le micro-ordinateur, la souris, l'interface graphique, le tableur électronique, ou le réseau local. Mais Xerox, dont ce n'était pas le métier, a laissé, involontairement, des entreprises comme Apple ou Microsoft profiter de ces nouveaux concepts. Dans le domaine des traitements de texte pour PC, MicroPro fut un pionnier avant d'être dépassé par WordPerfect puis Microsoft Word. Idem pour les tableurs, avec VisiCalc, dépassé par Lotus 1-2-3, puis Microsoft Excel. Idem encore avec les navigateurs Internet, domaine dans lequel Internet Explorer de Microsoft, encore, a supplanté Netscape. Etre le premier sur un marché, ne garantit pas le succès financier à long terme. Microsoft a ainsi réussi à s'imposer en copiant les autres, puis en apportant en permanence des améliorations incrémentales, et en submergeant le marché avec une politique de volume et des prix bas. Les produits Microsoft, généralement, arrivent à partir de la version trois, à se rapprocher des leaders du marché, en termes de fonctionnalité et de qualité pour les concurrencer efficacement. Ensuite, Microsoft s'impose en baissant fortement ses prix, ce qui élimine les concurrents sauf dans certains cas.

Un contre-exemple instructif est justement fourni par la société Intuit qui vend le logiciel Quicken et résiste à la concurrence de Microsoft Money, grâce à sa position de pionnier avec un très bon produit, qu'elle a amélioré sans cesse, depuis ses débuts en 1983. Intuit a survécu sans être racheté par un autre éditeur de logiciel, et s'est développé aussi par croissance externe, en achetant par exemple, d'autres produits grand public, comme TurboTax pour le calcul des

impôts. Comme les lois évoluent chaque année, les utilisateurs de ce dernier logiciel, sont des clients réguliers. Intuit a aussi fait des acquisitions dans le secteur connexe des services financiers. Son site Internet Quicken.com génère des revenus de publicité et de transactions de la part d'institutions financières présentes sur ce site.

De même, dans un autre domaine, Business Objects et son principal concurrent Cognos, sont devenus les leaders de la *Business Intelligence* (BI) ou informatique décisionnelle en étant les premiers et en proposant des solutions « sophistiquées ». Siebel a été rapidement le leader du marché du CRM. Netscape fut numéro un pendant une courte période des navigateurs et serveurs Internet. Il y a donc clairement un bénéfice à être un pionnier ou un leader technologique, car les premières entreprises qui investissent un marché ont quand même une bonne chance que leurs produits deviennent des standards ou des plates-formes, ce qui rend difficile le passage à un autre éditeur de logiciel, quand on est client.

### **Les *platform leaders* sont les acteurs centraux de l'industrie du logiciel :**

Le terme « plate-forme » fait référence, nous l'avons dit, à un produit « fondation » dont la valeur est maximale, quand il fonctionne comme le cœur d'un système de composants, développés par une ou plusieurs firmes. Les entreprises qui conçoivent et produisent de tels produits sont donc appelées *platform leaders* (Gawer et Cusumano, 2002 ; Gawer et Henderson, 2007). C'est le cas d'Intel, qui cible pour ses microprocesseurs et autres puces électroniques, aussi bien le marché des micro-ordinateurs fonctionnant avec Windows, que celui des nombreux matériels et logiciels complémentaires, de la caméra numérique au logiciel de gestion de finances personnelles. Microsoft adoptant une vision plus focalisée, considère que Windows est une plate-forme technologique pour les producteurs de logiciels d'application, et d'autres types de complément. Il y a eu et il y a toujours de nombreuses plates-formes qui dominent la concurrence en termes de part de marché, dans les marchés high-tech: les machines au standard VHS de Matsushita pour l'enregistrement vidéo ; les PDA tournant sous Palm OS, les serveurs d'application Web tournant sous Weblogic de BEA Systems ou WebSphere d'IBM.

Toutes les industries n'ont pas de plates-formes qui deviennent dominantes, mais beaucoup de marchés de masse high-tech et de niches de marché, gravitent vers un standard parce qu'elles dépendent de *core products* (ou produits cœur) et de produits complémentaires compatibles. On observe qu'il n'y a pas besoin d'être le premier sur le marché ou d'être un leader technologique

pour devenir *le platform leader* et atteindre une part de marché dominante. Le cas de Microsoft DOS et Windows l'illustre parfaitement : le système d'exploitation Mac OS était le meilleur au milieu des années 1980, de même que celui d'IBM, l'OS/2 au début des années 1990.

Les *platform leaders* doivent conduire l'innovation autour de leurs plates-formes particulières, au niveau de l'industrie, parce que leur valeur augmente à mesure que le nombre de compléments augmente. Plus les gens utilisent ce genre de produits de plate-forme, plus les incitations à introduire des produits complémentaires de la part des producteurs de compléments sont grandes ; ce qui pousse alors encore plus de gens à acheter ou utiliser la plate-forme, ce qui entraîne à nouveau plus d'innovations complémentaires, etc. On retrouve ici la notion d'« effet de réseau », déjà évoquée précédemment. Les *platform leaders* ont donc intérêt à favoriser l'innovation portée par les produits complémentaires, car ils n'ont généralement pas la possibilité de créer tous les compléments ou les systèmes de produits complets en interne. Ils doivent donc travailler de manière intime avec d'autres firmes, et combiner leurs efforts pour que la taille du marché de chacun augmente. Les *platform leaders* établissent de facto des standards dans leur industrie, et les éditeurs de progiciels doivent comprendre la dynamique de ces standards : elles doivent comprendre comment ces standards émergent pour augmenter les chances de succès commercial de leur innovation. Cusumano (2004) estime que le Macintosh n'est pas devenu un produit de masse parce que Apple n'a pas su devenir standard ou un *platform leader* : Apple a maintenu des prix élevés, n'a pas accordé de licences assez tôt, n'a pas fait en sorte qu'il y ait suffisamment de Macintosh sur le marché pour attirer autant d'acteurs de complément (dans le logiciel) que la concurrence, et n'a pas continué à innover dans le logiciel, ce qui a permis à Microsoft de le rattraper. Pour Cusumano, les managers d'Apple n'ont donc pas compris la dynamique des standards dans l'industrie du logiciel, ce qui a empêché le Macintosh de devenir un produit de masse, même s'il a réussi à s'imposer dans des niches de marché telles que la PAO (Publication Assistée par Ordinateur).

Dans l'ouvrage « *platform leadership* », Gawer et Cusumano (2002) ont identifié quatre leviers que les *platform leaders* peuvent utiliser pour exercer leur influence sur les producteurs ou les acteurs de complément :

- Le premier est le champ d'action de la firme, i.e. les compléments ou produits complémentaires, que la firme réalise elle-même. Ce n'est plus l'alternative « make or buy », car les *platform leaders* essayent généralement d'influencer d'autres firmes quant à leur décision de

produire des compléments valorisant la plate-forme, plutôt que d'acheter, comme dans le cas de l'intégration verticale. Dans tous les cas le *platform leader* doit déterminer s'il doit développer en interne ses propres compléments, ou laisser faire le marché, ou encore cultiver une petite capacité interne. Microsoft, par exemple, a encouragé beaucoup de tierces parties à développer des applications pour le DOS ou Windows : mais depuis le début des années 1990, la firme a les ressources pour concevoir et développer la plupart des compléments. Microsoft peut ainsi être quasiment certain que chaque génération de sa plate-forme (comme Windows XP ou Vista) sera au pire un succès modéré. D'autres « entreprises de plate-forme logicielle » ont du faire appel, dans une plus large mesure, à des tierces parties pour fournir des produits complémentaires.

- Le deuxième levier, est fourni par la technologie produit (architecture, interface et propriété intellectuelle). Les *platform leaders* doivent décider du degré de modularité de leurs architectures produit, et du degré d'ouverture des interfaces vers la plate-forme. Ils doivent aussi décider quelles informations sur la plate-forme et ses interfaces divulguer à des acteurs de complément potentiels, qui pourraient utiliser cette information pour devenir des concurrents directs. Pour le micro-ordinateur, Intel s'appuie sur les interfaces que de nombreuses entreprises ont été capables d'établir et de faire évoluer. Microsoft, par contre, a une position plus dominante avec Windows, et a établi des interfaces qui étaient principalement propriétaires.

- Le troisième levier est constitué par les relations avec les acteurs de complément. Les *platform leaders* doivent déterminer dans quelle mesure, ils seront leurs collaborateurs ou leurs concurrents. Ils doivent aussi penser à créer un consensus parmi leurs acteurs de complément, car des conflits d'intérêt peuvent émerger si certains envisagent d'entrer sur un marché où d'autres sont déjà présents. Microsoft a toujours limité l'étendue de son activité, tout en affirmant qu'il pourrait concurrencer ses acteurs de complément, si le marché était très prometteur. La stratégie de Microsoft est d'attaquer tout marché horizontal doté d'un gros potentiel, et sa stratégie pour Windows est d'empêcher la concurrence potentielle en améliorant régulièrement ce système d'exploitation grâce à des fonctionnalités intégrées, que des acteurs de complément vendent comme produits séparés.

- Le quatrième levier est constitué par l'organisation interne : les *platform leaders* doivent s'organiser pour être capables de gérer à la fois les conflits d'intérêt internes et externes. Ils peuvent décider de garder les groupes possédant des objectifs similaires dans une seule direction, ou alors de séparer ces groupes en plusieurs départements s'ils ont des objectifs potentiellement conflictuels. Ainsi, Intel a essayé de bâtir une barrière entre les groupes R&D et produits internes

qui pourraient avoir des intérêts conflictuels entre eux et des acteurs de complément externes. Par contre, Microsoft a longtemps affirmé qu'une telle séparation n'existait pas entre ses groupes « système d'exploitation » et ses groupes « application ». Microsoft insistait même sur le fait que l'intégration de différentes applications, systèmes, et technologies de réseau (comme l'intégration d'un navigateur Internet dans Windows, ou d'une messagerie instantanée dans Windows XP) était au centre de sa stratégie et profitable pour les clients. Microsoft avait la puissance suffisante pour pouvoir assumer ces choix, bien que les législations antitrust en Europe et aux Etats-Unis l'aient finalement contraint à revoir sa stratégie.

### **La plupart des éditeurs de logiciel sont des acteurs de complément :**

La majorité des entreprises de logiciel ne sont pas des *platform leaders*, mais des acteurs de complément. Cela veut dire qu'elles doivent prendre des décisions risquées quant à leur choix de plate-forme : faut-il écrire des applications pour Macintosh ou pour Windows ? Pour Linux ou pour une des versions d'Unix ? Doit-on développer pour l'environnement Java, ou pour son équivalent Microsoft, .NET ? Quels serveurs d'application Web (Weblogic, WebSphere, Apache, Windows 2000) doit-on utiliser comme environnement de ses applications ? Ce sont des décisions stratégiques, qui conditionnent l'activité de l'entreprise sur le long terme. Les choix de plate-forme sont liés aux autres choix stratégiques concernant les produits : choix entre marchés de masse (Windows, Palm OS) ou niches (Macintosh, Linux), choix entre marchés de grande consommation (Windows) ou d'entreprises (Unix, serveurs d'application Web).

Les principales difficultés pour un acteur de complément résultent de ce qu'il doit être capable de suivre les évolutions technologiques de la plate-forme sur laquelle il travaille, et qu'il risque de voir le *platform leader* absorber ses produits de compléments dans sa plate-forme en les produisant lui-même, comme Microsoft le fait régulièrement avec Windows, au travers d'acquisitions externes ou, plus souvent, de développements internes.

Par ailleurs, tenter de concurrencer les *platform leaders* peut être très dangereux pour les acteurs de complément, comme le montre l'exemple de Netscape, dont le navigateur Internet était un excellent complément de Windows. Mais quand Netscape a tenté de promouvoir son logiciel comme une plate-forme alternative à Windows, car les utilisateurs pouvaient lancer des applications Web à partir de ce navigateur plutôt que de Windows, Microsoft a réagi en fournissant gratuitement Internet Explorer avec ses nouvelles versions de Windows. Finalement,

Microsoft gagna face à Netscape, comme il l'avait déjà fait face à WordPerfect, Lotus 1-2-3, ou OS/2 d'IBM.

## **8. LE WEB SEMANTIQUE, ET L'EMERGENCE DES APPLICATIONS DE *TEXT MINING* :**

Nous allons dans la suite de notre travail, nous intéresser à la conception d'une innovation logicielle dans le domaine du *text mining*, domaine qui s'inscrit dans le cadre global du Web sémantique. Le *text mining* est un nouveau type d'application, un secteur innovant qui en tant que tel est constitué d'une multitude de petits acteurs innovants, mais auquel depuis deux ans, les grandes entreprises s'intéressent, ce qui s'est traduit récemment par quelques acquisitions spectaculaires. Nous allons donc parler successivement du web sémantique, du *text mining* et de ses applications, des opérations de consolidation dans ce secteur, et de la stratégie d'IBM sur ce marché.

### **Le Web sémantique :**

Le Web sémantique, tel que le W3C (World Wide Web Consortium, qui promeut la compatibilité des technologies du web) l'a défini, consiste en un ensemble de technologies visant à rendre le contenu des ressources du World Wide Web accessible et utilisable par les agents logiciels, grâce à un système de métadonnées formelles. Pour un certain nombre de spécialistes d'Internet, le Web sémantique constitue la prochaine grande évolution des réseaux basés sur le langage XML et le protocole IP. L'objectif fondamental du Web sémantique est que l'ensemble des informations publiées en ligne soit reliées entre elles, de façon à créer la plus vaste base de connaissance du monde. Tandis que le World Wide Web a été conçu pour les humains, le Web sémantique vise à doter l'ensemble des ressources en ligne d'un sens, interprétable par un logiciel. Grâce au Web sémantique, les moteurs de recherche n'auront plus à interpréter les pages HTML et à tenter d'y découvrir des mots clés. Chaque document, de la page XHTML à la fiche produit basée sur XML, se décrira à l'aide d'un vocabulaire normé, reposant sur une ontologie (i.e. un modèle de données représentatif des concepts d'un domaine et des relations entre ces concepts). Le Web sémantique s'appuie sur les standards RDF et OWL, définis par le W3C. Le premier fournit le mécanisme technique pour représenter des relations entre ressources. Le second s'appuie sur RDF pour guider les entreprises dans la construction d'ontologies.

Pour certains spécialistes et acteurs du monde Internet, l'accès à l'information est la prochaine "nouvelle frontière", le territoire à conquérir pour assurer la performance de l'entreprise et de ses collaborateurs dans un contexte où tout s'accélère. Par ailleurs, les risques économiques, politiques et sociaux qui pèsent sur les entreprises sont de plus en plus grands. Cette pression demande de nouveaux outils de gestion de l'information, d'accès à la connaissance, de partage et de transmission des savoirs. C'est dans ce contexte qu'ils espèrent l'avènement du Web sémantique.

### **Le *text mining*:**

La fouille de textes ou extraction de connaissances dans les textes est une spécialisation de la fouille de données et fait partie du domaine de l'intelligence artificielle. Cette technique est généralement désignée sous l'anglicisme *text mining*. C'est un ensemble de traitements informatiques consistant à extraire des connaissances selon un critère de nouveauté ou de similarité dans des textes produits par des humains pour des humains. Dans la pratique, cela revient à mettre en algorithmes un modèle simplifié des théories linguistiques dans des systèmes informatiques d'apprentissage et de statistiques. Les disciplines impliquées sont donc la linguistique calculatoire, l'ingénierie du langage, l'apprentissage artificiel, les statistiques et bien sûr l'informatique. Le domaine du *text mining* reste à l'heure actuelle un domaine nouveau, qui s'inscrit dans le cadre des technologies du Web sémantique, et dont la maturité est insuffisante pour que les éditeurs de logiciel qui l'ont investi atteignent des niveaux de rentabilité élevés.

Un des premiers secteurs marchands à s'être approprié la technologie du *text mining*, et ce depuis quelques années, est celui de l'industrie pharmaceutique, dont les services de recherche et de développement utilisent le *text mining* pour faire de la veille scientifique. Mais les laboratoires utilisent également ces technologies à des fins de marketing. Lors du lancement d'une nouvelle molécule, il peut être intéressant d'étudier les réactions des clients à travers l'étude de blogs. Le laboratoire Pfizer a, par exemple, investi dans une solution de veille de l'éditeur Temis, chargée de surveiller les blogs de patients : si un blogueur évoque des effets secondaires inédits suite à la prise d'une nouvelle molécule, l'application envoie une alerte. Autres habitués du *text mining*, les fournisseurs d'informations payantes, Factiva ou Thomson, qui intègrent progressivement ces technologies dans leurs offres.

Par ailleurs, les applications du *text mining* sont potentiellement très nombreuses. L'une d'entre elles est la gestion de la relation client : prenons l'exemple des réclamations rédigées à l'attention d'un opérateur de télécommunications par ses clients : « *Ma connexion à internet est rompue depuis des semaines, j'ai essayé de vous joindre vingt fois, en vain, je résilie mon contrat.* » Ce message posté sur le site d'un opérateur ne sera pas exploité à des fins d'analyse. En effet, une telle analyse ne s'applique actuellement qu'aux données structurées stockées dans un entrepôt (ou *datawarehouse*) : référence client, nom, type d'équipement... Les entreprises gagneraient pourtant à ordonner et exploiter ce gisement d'informations. Un outil de *text mining*, permet d'extraire l'information de documents non structurés, comme des textes officiels, documents, messages, ou autres articles de presse, par exemple. Les outils de *text-mining* ont donc pour vocation d'**automatiser la structuration des documents** peu ou faiblement structurés. Ainsi, à partir d'un document texte, un outil de *text mining* va **générer de l'information sur le contenu du document**. Cette information n'était pas présente, ou explicite, dans le document sous sa forme initiale, elle va être rajoutée, et donc enrichir le document. Cela peut servir à classifier automatiquement des documents, à avoir un aperçu du contenu d'un document sans le lire, à alimenter automatiquement des bases de données ou de connaissance, à faire de la veille sur des corpus documentaires importants, ou encore à enrichir l'index d'un moteur de recherche pour améliorer la consultation des documents.

### **L'intérêt des grandes entreprises et la consolidation du secteur :**

Le secteur du *text mining* était jusqu'à une période récente, uniquement occupé par de petits éditeurs de logiciel, dont les produits ont été conçus à partir de travaux de recherche menés au sein de centres de recherche publics ou privés. Ainsi, les recherches conduites dans des centres de recherche américains (Xerox PARC) et européens (XRCE) de Xerox, ont abouti après essaimage à la création de deux sociétés concurrentes, l'une InXight basée aux Etats-Unis qui fournit son logiciel à de nombreuses agences gouvernementales de renseignements ou à des services de l'armée pour faire de la veille et traquer leurs cibles en analysant les textes publiés sur le réseau, l'autre Temis en France qui s'est initialement orientée vers le domaine de l'intelligence économique et de la veille technologique pour les grands laboratoires pharmaceutiques. Ces deux acteurs de référence du marché des logiciels de *text mining*, sont loin d'être seuls, et de nombreuses entreprises concurrentes ont ainsi vu le jour depuis le début des années 2000, chacune prétendant détenir un positionnement original par rapport à l'ensemble de ses concurrentes, ce qui n'est que partiellement vrai.

Mais si les technologies gagnent en maturité, les clients sont encore assez rares. Chacun de ces éditeurs n'en compte au mieux que quelques dizaines, car la priorité pour les entreprises clientes porte encore sur l'analyse des données stockées dans des bases. Néanmoins le potentiel commercial de ses logiciels d'application a fini par attirer la convoitise de grands acteurs du marché du logiciel, persuadés que l'analyse textuelle allait devenir une technologie incontournable, et que le capital de connaissances d'une entreprise se situait plus dans les documents textuels que dans les champs d'une base de données. Cela a entraîné depuis 2005, une série de rachats s'inscrivant dans un contexte de plus en plus évident de consolidation du secteur. Ainsi en témoignent les récents rachats d'InXight (120 personnes et 25 millions de dollars de chiffre d'affaires en 2006), de ClearForest, et de Fast. Désireux d'investir le marché de l'analyse et de la recherche de données non structurées, Business Objects a racheté InXight en mai 2007, Reuters a acquis ClearForest toujours en 2007, et Microsoft s'est emparé de Fast en 2008, pour 1,2 milliards de dollars. En ce qui concerne Business Objects, l'éditeur de logiciel franco-américain spécialisé dans le décisionnel, veut intégrer la solution d'analyse, de recherche, et de visualisation de données non structurées conçue par InXight, qui va enrichir sa plate forme de business intelligence Business Objects XI : *« La recherche assistée, l'analyse de texte et l'accès aux données non structurées font partie intégrante de la business intelligence (mais) représentent une source de connaissances dont la plupart des entreprises ne cherchent pas encore à profiter. Or, qu'elle ait pour but la détection des fraudes, le respect de la réglementation, la connaissance des clients ou le renseignement et la lutte contre le terrorisme, l'analyse de texte peut aider à prévoir des événements importants, à détecter des opportunités »*, a souligné John Schwarz, directeur général de Business Objects, en cette occasion. Microsoft, pour sa part, désire ne pas se laisser distancer et a préféré réagir vite en rachetant le leader du secteur, Fast, sans pour autant avoir la garantie de réaliser une solution résolvant effectivement les besoins de recherche des grandes entreprises, car la digestion de cette acquisition sur le plan commercial, technique et opérationnel peut prendre du temps.

### **IBM : un acteur majeur, une stratégie différente :**

Par rapport à autres poids lourds de l'industrie logicielle, IBM a choisi une stratégie différente. Il faut dire que la firme est présente sur le marché depuis plusieurs années avec son *framework* UIMA, une plate-forme susceptible d'intégrer tous les logiciels moteurs d'analyse (dont les logiciels de *text mining*) qui choisiront d'être compatibles avec ce standard, et avec son moteur de recherche OmniFind, qui est capable d'exploiter les métadonnées fournies par ces différents logiciel (UIMA signifiant *Unstructured Information Management Application*, i.e. application de gestion

d'information non structurée). IBM a quand même lui aussi procédé à des rachats, mais d'acteurs de plus petite taille, comme iPhrase en novembre 2005. Grâce à cela, le moteur de recherche OmniFind exploite les données provenant du module Content Discovery Server, moteur d'interprétation d'iPhrase qui extrait des informations contextuelles relatives aux requêtes des utilisateurs.

L'architecture UIMA peut être vue comme un bus composé d'une série de moteurs d'analyse et alimenté en entrée par un document ou une collection de documents. Lorsque le premier moteur achève son traitement, il conserve le résultat de ses analyses dans un système commun d'analyse (ou CAS) sous forme de métadonnées. Le CAS est défini par un modèle de données et un index, qui rend accessible au second moteur d'analyse les métadonnées générées par le premier. Quant aux moteurs (les composants UIMA), ils peuvent être développés par n'importe quel adhérent à UIMA (éditeurs, universitaires, scientifique...). *Le framework* ou environnement UIMA vise donc à favoriser l'interopérabilité des résultats d'analyse de données non structurées : *text mining*, catégorisation, moteur de recherche, indexation audio, vidéo... Jusqu'à présent, les métadonnées récoltées à l'issue des traitements restaient exprimées dans un format spécifique au logiciel. Difficile alors pour le moteur d'analyse d'un éditeur de récupérer et de traiter à son tour les résultats d'un autre éditeur. C'est pour pallier cette limite du monde non structuré qu'IBM a lancé et « ouvert » UIMA. Ce *framework* de développement gère notamment l'interopérabilité des métadonnées entre les moteurs d'analyse, et en 2007 une quinzaine d'acteurs s'y étaient ralliés, dont ClearForest, Cognos, Factiva, InXight, SPSS, ou Temis. Ainsi la solution de text-mining de Temis, Luxid, est « UIMA compliant », c'est-à-dire qu'elle est compatible avec ce standard, et qu'elle peut être « pluggée », i.e. connectée, au bus UIMA d'IBM.

A l'issue des traitements successifs, UIMA compose des CAS « prêts à l'emploi », dont les données structurées sont envoyées dans des bases relationnelles, ou dans des moteurs de recherche pour enrichir leur propre index. C'est d'ailleurs ce que fera l'outil de recherche d'IBM : OmniFind dépassera ainsi l'approche par mots clés pour s'ouvrir à la sémantique via les moteurs d'analyse de texte.

En conclusion, à propos d'IBM, comme UIMA est une architecture ouverte et libre, le géant américain souhaite s'imposer, non pas comme éditeur de logiciel de *text mining*, mais comme développeur de l'architecture qui permet d'accueillir divers outils de *text mining* tiers. La question se pose de savoir si, au bout de la chaîne UIMA, le moteur de recherche qui va exploiter toute

cette chaîne de traitement peut être autre chose que le moteur OmniFind d'IBM. En somme, IBM, acteur reconnu du *middleware* (logiciels situés entre les couches basses, comme le système d'exploitation, et les couches hautes, logiciels d'application, d'un système informatique), a choisi la stratégie qui consiste à se positionner comme le *middleware* ouvert du monde du traitement des informations non-structurées. Mais pour qu'UIMA devienne le standard du non structuré, SAP, Microsoft, Oracle, Verity, Autonomy devront s'y rallier.

Le secteur du *text mining* rentre donc dans une phase de consolidation, phase au cours de laquelle des standards devraient émerger, ainsi que de véritables leaders, susceptibles à la fois de développer le marché et d'y occuper une position dominante.

**Chapitre 2 :  
Apports de la littérature : de la  
modularisation à l'innovation  
par les utilisateurs**

## SOMMAIRE DU CHAPITRE 2 :

<b>Chapitre 2 : Apports de la littérature : de la modularisation à l'innovation par les utilisateurs</b>	<b>42</b>
<i>Sommaire du Chapitre 2 :</i>	43
<b>1. La littérature relative à la notion de Modularité :</b>	<b>46</b>
1.1 Introduction :	46
1.2 Définitions :	46
1.3 Les principaux thèmes abordés par la littérature sur la modularité :	50
1.4 La question de l'émergence des <i>design rules</i> :	54
1.5. Deux exemples emblématiques dans l'industrie informatique : les ordinateurs IBM s/360 et le système d'exploitation Unix.	55
1.5.1 Les ordinateurs IBM s/360 :	55
1.5.2 Le système d'exploitation Unix :	57
1.6. Les conséquences de la modularité sur les organisations et la structure d'une industrie :	58
1.6.1 Réseaux de fournisseurs de systèmes modulaires :	58
1.6.2 Architecte ou modulier, deux stratégies exclusives ?	60
1.6.3 Correspondance entre architecture produit et structure organisationnelle :	61
1.7. Modularité et innovation :	62
1.7.1 Introduction :	62
1.7.2 L'innovation relative aux produits modulaires :	63
1.7.3 Architecture produit et exploration des marchés :	65
1.7.4 Limites de l'architecture modulaire :	66
1.7.5 Architecture et processus de conception :	66
1.7.6 Modularité et innovation : conclusion	68
1.8 Conclusion et questions soulevées par la littérature sur la modularité :	68
<b>2. Les apports de la littérature sur l'innovation par les utilisateurs : le logiciel comme champ d'application idéal ?</b>	<b>70</b>
2.1 Les premiers travaux de von Hippel :	71
2.2 La question de la source de l'innovation : comment les entreprises captent-elles les rentes liées à l'innovation ?	72
2.3 Le concept de <i>lead user</i> :	74
2.4 L'adhérence (« <i>stickiness</i> ») de l'information :	77
2.5 Lien entre la théorie de la diffusion et le concept de <i>lead user</i> :	78
2.6 Le concept de <i>lead user</i> comparé à d'autres concepts de la littérature :	79
2.7 Le concept de <i>lead user</i> dans la littérature :	79
2.8 Une approche « universelle » ?	81
2.9 L'approche <i>lead user</i> pour générer des innovations radicales :	82
2.10 La notion de <i>toolkit</i> : définition et lien avec l'innovation utilisateur	84
2.10.1 Introduction :	84
2.10.3 Le développement des <i>toolkits</i> par le fournisseur :	86
2.11 La contribution des communautés d'utilisateurs :	87
2.12 Conclusion sur l'approche <i>user innovation</i> :	89

3.	<i>La perspective relationnelle retenue par le marketing interorganisationnel :</i>	89
3.1	Introduction :	89
3.2	Les travaux du groupe IMP :	90
3.2.1	La perspective relationnelle :	90
3.2.2	De la dyade au réseau :	91
3.2.3	Le modèle d'interaction :	91
3.2.4	L'innovation comme produit de l'interaction client/fournisseur :	93
3.3	Conclusion :	97
4.	<i>Conclusion sur les apports de la littérature :</i>	98

Nous cherchons à répondre à la question de recherche suivante : « quels rôles jouent les premiers clients dans le processus de conception et comment les associer à ce processus ? ». Nous avons choisi d'étudier cette question dans le secteur des logiciels destinés aux entreprises. Cela nous a conduits à retenir les apports de différents champs de la littérature. Tout d'abord, la littérature portant sur la notion de modularité est apparue incontournable parce que le concept de modularité est prégnant dans la conception de logiciel. Depuis que les résultats des travaux de recherche de Parnas (1972), ont été exploités et mis en application par les praticiens de ce secteur, le logiciel est en effet un des champs d'application privilégié de ce concept. Cela nous conduira à aborder les notions de *design rules* des produits dotés d'architecture modulaire, et au processus qui conduit à leur émergence. Nous nous intéresserons ensuite aux enjeux organisationnels de la modularité, et à son impact sur la structuration d'une industrie. La modularisation des architectures logicielles peut permettre l'éclatement du processus de conception entre une multitude d'acteurs. Ainsi, certains utilisateurs, suffisamment compétents, peuvent personnaliser des logiciels, grâce à des modules particuliers, les *user toolkits* ou environnements de développement logiciel destinés aux utilisateurs. Cela illustre le lien entre la modularité d'une architecture produit et la capacité des utilisateurs à développer leurs propres produits ou, tout du moins, leurs propres personnalisations. Justement dans le secteur du logiciel, l'innovation conçue par les utilisateurs est un phénomène majeur tant au niveau des logiciels Open Source que des logiciels propriétaires (von Hippel, 2001). Ainsi notre question de recherche, centrée sur le rôle des clients dans la conception logicielle, nous amène-t-elle logiquement à analyser les apports de ce champ, qui tend à considérer les utilisateurs, comme les acteurs les mieux placés pour concevoir et développer les innovations logicielles. Enfin, nous prendrons en compte les apports de la littérature dans le champ du marketing industriel (ou marketing interorganisationnel), qui, comme la littérature sur l'innovation par les utilisateurs remet en cause le paradigme « manufacturer active » selon lequel les innovations radicales sont conçues par les fournisseurs seuls. Par contre, les travaux dans le champ du marketing industriel, notamment ceux menés par les chercheurs du groupe IMP (Hakansson, 1982), soulignent le caractère multidimensionnel et suivi des relations entre fournisseur et client, ce en quoi ils se distinguent des travaux issus du champ de l'innovation par les utilisateurs, qui adopte une vision plus épisodique et monodimensionnelle de ces relations.

## 1. LA LITTÉRATURE RELATIVE A LA NOTION DE MODULARITE :

### 1.1 Introduction :

La notion de modularité a été mise en œuvre dans un certain nombre de domaines, et a généré une littérature académique relativement abondante, surtout à partir du milieu des années 1990. Les raisons du succès du concept de modularité résulteraient principalement des pressions économiques considérables qui poussent vers une division du travail entre les concepteurs de produits et les fournisseurs, sur la base de composants modulaires normalisés (Ulrich 1995). Brusoni et Prencipe (2002) pensent que la modularité et la désintégration organisationnelle sont les réponses cognitives et organisationnelles les plus récentes à l'augmentation de la complexité et à la spécialisation systémiques et cognitives. Fixson (2003) observe qu'il n'y a pas de consensus précis quant à la définition exacte de cette notion. Il observe la modularité à différents stades du processus de création de produit : « la modularité est utilisée comme un moyen d'atteindre différents objectifs à différentes étapes du cycle de vie des produits ». L'auteur en conclut que cela complique fortement les comparaisons entre les différentes descriptions de la modularité, et que cela explique pourquoi les chercheurs ont conduit peu de travaux empiriques sur ce thème.

### 1.2 Définitions :

Nous aborderons donc cette notion de modularité en précisant les définitions que l'on trouve dans la littérature et celles que nous retiendrons. A l'origine, la connaissance pratique de la modularité provient pour une grande part de l'industrie informatique. Le terme « architecture » a été utilisé pour la première fois à propos des ordinateurs, par les concepteurs des IBM 360 (Amdahl, Blaauw et Brooks, 1964), tandis que la base scientifique de l'architecture des ordinateurs a été établie par Bell et Newell (1971). La littérature relative à la modularité, a aussi été rapidement féconde dans le champ du logiciel, car dès 1972, Parnas a établi des principes tels que celui d'« *information hiding* », ou masquage d'information, qui ont été largement repris par la littérature. Le terme « *design rules* », désignant les règles de conception d'une architecture modulaire, fut quant à lui utilisé en premier lieu par Mead et Conway en 1980, tandis que les innovations architecturales de Sun ont, pour leur part, été décrites par Hennessy et Patterson en 1990.

Comme souligné en introduction, cette notion a donné lieu à de nombreux travaux, notamment à partir des années 1990. Cependant comme le note Fixson (2003), il y a un grand nombre de façons de décrire la modularité : « un module contient une fonction technique » ; « l'utilisateur est capable de reconfigurer les modules et donc le produit » ; « la modularité permet de réduire la complexité lors de la phase d'assemblage ».

Pour Baldwin et Clark (2000), le concept de modularité inclut **deux idées** :

- **la première** est celle d'**interdépendance à l'intérieur de chaque module et d'indépendance entre chaque module**. Un module est défini quant à lui, par une unité dont les éléments structurels sont puissamment connectés entre eux et relativement faiblement connectés aux éléments des autres unités. Il y a ainsi des degrés de connectivité, et donc des degrés de modularité (définition adaptée de McClelland et Rumelhart, 1995, par Baldwin et Clark, 2000). Cette définition diffère de celle donnée par Ulrich (1995) pour qui « une architecture (produit) modulaire est une architecture qui inclut un isomorphisme (*mapping one to one*) entre les éléments fonctionnels et les composants physiques du produit, et spécifie les interfaces découplées entre composants ». Pour Baldwin et Clark (2000), cette définition de la modularité basée sur les fonctions n'est pas complètement satisfaisante car les fonctions sont par essence multiples et changeantes, alors que leur propre définition est basée sur des relations entre structures. Quoiqu'il en soit, les différents auteurs s'accordent sur le fait que les modules sont des unités d'un système plus large, qui travaillent ensemble tout en étant structurellement indépendantes les unes des autres.

- **la deuxième idée**, est portée par les termes ; ***abstraction, information hiding* ou masquage d'information**, d'une part, et **interface**, d'autre part. Les auteurs proposent à ce propos la synthèse de la littérature suivante : « un système complexe peut être géré en le divisant en plus petits morceaux et en regardant chacun séparément. Quand la complexité d'un élément dépasse un certain seuil, cette complexité peut être isolée en définissant une *abstraction* séparée qui a une interface simple. L'*abstraction* masque la complexité de l'élément ; l'*interface* indique comment l'élément interagit avec le système. » (Baldwin et Clark, 2000 ; p. 64).

Pour Sanchez et Mahoney (1996), la modularité est une forme particulière de concept qui crée un haut degré d'indépendance ou « faible couplage » entre des concepts de composants en standardisant les spécifications d'interface entre composants (un faible couplage entre composants dans un concept de produit est différent d'un faible couplage dans un produit réel,

généralement physique : il en est ainsi par exemple des composants faiblement couplés d'un micro-ordinateur, mais fortement couplés physiquement). Les relations entre composants sont définies par les spécifications des *inputs* et *outputs* reliant les composants dans un concept, et un ensemble complet de spécifications d'interface constituent une *architecture produit* (Abernathy et Clark, 1985 ; Clark, 1985).

Pour Baldwin et Clark (1997), la modularité est une stratégie visant à organiser efficacement les processus et les produits complexes. Un système modulaire est composé d'unités (ou modules) qui sont conçues indépendamment mais fonctionne comme un ensemble intégré. Les concepteurs aboutissent à la modularité en découpant l'information en règles de conception visibles et en paramètres de conception cachés. La modularité n'est profitable que si la partition est précise, non ambiguë, et complète. Les règles de conception visibles (appelées aussi « information visible ») sont établies très tôt dans le processus de conception et communiquées largement à ceux qui sont impliqués. Les règles de conception (ou « *design rules* ») visibles appartiennent à l'une des trois catégories suivantes :

- Une architecture, qui spécifie quel module fera partie du système et quelles fonctions existeront ;
- Des interfaces qui décrivent en détail comment les modules vont interagir, i.e. comment ils se correspondront l'un l'autre, comment ils se connecteront, comment ils communiqueront ;
- Des standards permettant de tester la conformité d'un module aux *design rules* ou règles de conception (le module X peut-il fonctionner dans le système ?), et de mesurer la performance relative d'un module par rapport à un autre.

Les praticiens regroupent parfois les trois éléments de l'information visible et les appellent simplement « l'architecture », « les interfaces » ou « les standards ». Les paramètres de conception cachés (appelés aussi « information cachée ») sont des décisions qui n'affectent pas la conception au-delà du module local. Les éléments cachés peuvent être choisis tardivement, souvent modifiés, et n'ont pas à être communiqués à quiconque au-delà de l'équipe de conception du module (Baldwin et Clark, 1997).

## Définitions dans le domaine du logiciel :

Dans le secteur du logiciel, des travaux spécifiques, et précurseurs, ont été conduits sur cette question, à la suite de l'article séminal de Parnas (1972). L'auteur a en effet tenté dans un article qui a connu un écho très important dans l'industrie du logiciel et au-delà, de comprendre pourquoi les recommandations faites par les représentants académiques n'étaient à l'époque pas reprises dans le monde des développeurs de logiciel, alors que la qualité des logiciels développés par ces derniers était notoirement insatisfaisante. Pour cela, il choisit de dupliquer un programme « temps réel » destiné à l'aviation, l'*Onboard Flight Program* (OFP). Ce programme arrivait très difficilement à fonctionner en temps réel, entre autres problèmes, c'est pourquoi Parnas le choisit comme champ d'application pour les « idées nouvelles » prônées par les chercheurs. L'approche modulaire retenue par Parnas et son équipe a permis d'obtenir de tels résultats que les praticiens du secteur en ont modifié leur façon de travailler, par la suite : un des principes de base de cette approche, est la décomposition d'un projet logiciel en modules, ou tâches (« *work assignments* »). A l'époque, aucun projet de grande taille ne reprenait cette idée, largement promue par le monde académique. Ainsi, Parnas (1972) affirme que certains concepts complémentaires doivent accompagner la mise en œuvre de ce concept.

Parnas (1972) a défini trois structures importantes pour les programmeurs concernés, dans le cas du logiciel : la structure modulaire, la structure des usages, et la structure des processus.

1. Un module est une tâche à réaliser par un programmeur ou une équipe de programmeurs. Chaque module consiste en un groupe de programmes intimement liés. La structure modulaire (ou « *module structure* ») est constituée de la décomposition du programme en modules, et des hypothèses que chaque équipe responsable d'un module peut faire sur les autres modules.

2. Dans la structure des usages, les composants sont des programmes, i.e. des parties de modules et non des modules.

3. La structure des usages détermine les sous-parties exécutables du logiciel. La structure des processus est la décomposition des activités *run-time* du système en unités appelés processus. Les processus ne sont pas des programmes ; il n'y a pas de lien direct entre modules et processus, et chaque processus peut invoquer des programmes dans plusieurs modules.

L'article de Parnas (1972) traite exclusivement du premier point, la structure modulaire du logiciel. L'auteur définit la *structure en module*, comme étant fondée sur le critère de décomposition appelé « *information hiding* » (IH), ou masquage de l'information. Selon, ce principe, les détails du système qui risquent de changer indépendamment doivent rester les secrets de chaque module, et les hypothèses relatives aux interfaces entre modules ne doivent pas changer. Chaque structure de données (i.e. structure logique destinée à contenir des données) est utilisée dans un seul module : on peut y accéder par un ou plusieurs programmes à l'intérieur du module, mais pas par des programmes à l'extérieur du module. Tout autre programme qui requiert des informations stockées dans les structures de données d'un module, doit l'obtenir en appelant le programme d'accès appartenant à ce module. Il doit être possible de changer la mise en œuvre d'un module sans connaître celle des autres modules et sans affecter le comportement des autres modules :

- Il doit être possible de faire des changements dans un module, sans changer aucune interface de module,
- Il doit être possible de faire des changements logiciels majeurs, au travers de changements dans différents modules individuels, sans que les programmeurs modifiant un module aient besoin de communiquer avec les autres. Si les interfaces des modules ne sont pas modifiées, il doit être possible de faire tourner et de tester n'importe quelle combinaison de modules anciens ou nouveaux.

Les principes établis par Parnas (1972) ont depuis fait école dans l'univers du logiciel, où la modularité des architectures est depuis devenue la règle. Appliquer ces principes n'est pas toujours simple, c'est une tentative de minimiser le coût des logiciels et cela nécessite que le concepteur sache évaluer la probabilité que surviennent différents changements. De telles estimations, sont basées sur l'expérience du passé, et peut demander une connaissance du domaine d'application, aussi bien que la compréhension de la technologie logicielle et matériel.

### **1.3 Les principaux thèmes abordés par la littérature sur la modularité :**

Nous allons présenter ici de manière synthétique les principaux thèmes abordés par la littérature, puis nous nous focaliserons, lors des points suivants, sur certaines thématiques particulières. Au total, la littérature traitant de la modularité aborde des thèmes variés et situés à différents niveaux, sur la base d'exemples issus très largement de l'univers informatique :

1. Ainsi un thème a trait aux **conditions préalables à la modularisation** : elles sont, selon la littérature, les suivantes :

1. La capacité à spécifier des règles de conception *ex ante*,
2. La capacité à créer un jeu d'interfaces permettant des ajouts « *plug and play* » au niveau du processus d'assemblage,
3. La capacité à créer une organisation qui reflète les deux premières conditions.

Les deux premières contraintes émanent des bureaux d'études, alors que la troisième émane du management stratégique et de l'organisation (Baldwin et Clark, 1997). Dans la très grande majorité des cas, la littérature décrit le processus de modularisation, comme un processus de « dé-intégration » d'un produit doté d'une architecture intégrée. L'étude d'une trajectoire de ce type amène les chercheurs à se préoccuper des conditions préalables à la définition d'une architecture modulaire, et à définir les *design rules* ou règles de conception idoines.

2. Un second thème renvoie à la question des **bénéfices liés à la modularisation**, dans les processus de conception et de production. En ce qui concerne la conception, les principaux auteurs (Baldwin et Clark, 1997 et 2000 ; Ulrich, 1995 ; Langlois et Robertson, 1992 ; Henderson et Clark, 1990 ; Schilling, 2000 ; Parnas, 1972) affirment que la modularisation des architectures conduit à une optimisation du processus de conception de produit, et une augmentation du taux d'innovation des entreprises qui l'adoptent. Cela résulte principalement du fait que dans le cadre d'une architecture modulaire, le processus de conception peut être subdivisé en sous-tâches indépendantes les unes des autres, réalisables simultanément : le couplage faible entre modules permet alors l'innovation par des organisations indépendantes, focalisées sur un module particulier. L'innovation peut également provenir de la possibilité de mettre en place des conceptions d'éléments modulaires en simultané et en parallèle, ce qui permet là encore de réduire considérablement les délais d'exécution relatifs à un changement de modèle, en d'autres termes, une innovation systémique (Aoki et Takizawa 2002). Ce phénomène a été observé dans l'industrie automobile japonaise, et rappelle la conception de Baldwin et de Clark (1997) sur la modularité et la possibilité de mise en œuvre simultanée des différentes parties d'un plan de conception de grande envergure. Toutefois, la mise en œuvre du concept de modularité dans la construction automobile, concerne avant tout des sous-ensembles localisés, et ne permet pas de parler de modularité de conception, contrairement à l'industrie informatique. Langlois et Robertson (1992), quant à eux, mettent en exergue le fait qu'une architecture produit modulaire

favorise l'exploration d'un marché, le «*real time market research*» (Sanchez et Sudharashan, 1993) et le «*fine tuning*» d'un nouveau produit, comme cela a pu être constaté dans des domaines aussi différents que la construction automobile (Cusumano et Nobeoka, 1998 ; Fourcade, 2004) ou l'édition de logiciel (Korson et Vaishvani, 1986). Des divergences apparaissent toutefois entre les auteurs en ce qui concerne l'innovation architecturale, la plupart d'entre eux, tenant de la primauté de la modularité des architectures produit, estiment qu'elle est favorisée par le découplage entre innovation architecturale et innovation au niveau des modules (Sanchez et Mahoney, 1996). A l'inverse d'autres auteurs (Ulrich, 1995) estiment que les concepteurs des modules étant isolés, concentrés chacun sur leur module, l'innovation « architecturale » (Henderson & Clark 1990) est rendue presque impossible, privant la firme de cette option d'innovation. Dans ce cas, la conception s'enferme dans une « dépendance de sentier » («*path dependency*», Chesbrough & Kusunoki 1999/2001) dont il sera difficile de sortir.

En ce qui concerne la production, les bénéfices de la modularité proviennent de la possibilité de produire un grand nombre de variantes de produit à moindre coût, sur la base d'une même architecture comme le notait déjà Starr (1965). Pour Ulrich (1995), la flexibilité en production favorisée par les architectures modulaires permet la différenciation au dernier moment du produit final, alors que Gilmore et Pine (1993) estiment qu'elle rend possible la personnalisation de masse («*mass customization*»). Pine (1999) a montré que les turbulences de marché et le besoin en variétés de produit ont cru nettement, et affirme que ce phénomène va se poursuivre. La difficulté consiste alors à créer de la variété à un coût faible. Upton (1997) définit alors la modularité comme « la capacité à changer ou adapter avec peu d'efforts, de temps, ou de pénalité ». De même, Ulrich (1995), affirme que la capacité d'un système de production à créer des variétés réside non pas dans la flexibilité de l'équipement de production, mais dans l'architecture du produit. La flexibilité de cet équipement et l'architecture produit interagissent pour contribuer à la capacité de créer économiquement la variété produit. La variété peut être obtenue par combinaisons de *building blocks*. La modularité du produit permet de créer la variété à l'assemblage final. Avec une architecture produit modulaire, la variété peut être obtenue avec et sans équipement de production flexible (alors que pour l'architecture intégrée, un tel équipement est indispensable).

3. La littérature aborde également des **aspects organisationnels** liés au choix d'une stratégie de modularisation des produits. La plupart des chercheurs semblent en effet penser qu'il existe un lien entre une architecture de produit et une structure organisationnelle. Pour Sanchez et

Mahoney (1996), il est même possible que « si des organisations conçoivent des produits, des produits conçoivent des organisations », car « les tâches de coordination dans des conceptions de produit déterminent largement les concepts d'organisation permettant de développer et de produire ces produits ». De même, Baldwin et Clark (1997) affirment que « pour se battre dans un monde de modularité, les leaders doivent reconcevoir leur organisation interne [...] la réponse [...] est la modularité dans l'organisation ». Il s'agit ici d'éclater le travail, sur les différents modules constitutifs du produit, entre des équipes indépendantes. Pourtant, pour d'autres auteurs, il reste difficile de percevoir de quelle façon la modularisation en conception et en production peuvent s'apparenter à la conception organisationnelle intra-firme et inter firmes. En effet, « [...] la présomption d'un lien entre la modularité d'un produit et la modularité organisationnelle ne repose que sur quelques enquêtes empiriques » (Schilling et Steensma, 2001). Tout comme Fixson (2003) qui affirme que l'isomorphisme entre architecture produit et architecture organisationnelle reste à démontrer.

4. La modularisation d'une architecture produit peut avoir pour conséquence **la transformation d'une industrie**. Ainsi, l'adoption d'une stratégie axée sur la modularité des architectures produit peut conduire à un éclatement de l'industrie, et à l'apparition d'un grand nombre d'acteurs, qui à leur tour prennent des décisions stratégiques, notamment en ce qui concerne l'innovation. Les sources d'innovation sont alors multiples, ce qui est justement très favorable à la conception d'innovations, en particulier dans des conditions de marché et dans des environnements technologiques incertains. Il est ainsi possible de couvrir un vaste espace produit, et donc de réaliser une exploration marché performante. Dans cette exploration, certains clients jouent même un rôle crucial, qui peuvent par exemple créer des produits virtuels en assemblant des produits fournis par différents fournisseurs. Ce phénomène cité par Ulrich (1995) a toutefois été peu traité dans la littérature, et ne l'a été que dans le cas où des modules parfaitement définis pouvaient s'intégrer à une solution grâce à des interfaces elles-mêmes parfaitement définies, on parle alors de *modularité d'usage*. Plus généralement, Sanchez et Mahoney (1996) estiment que le découplage des processus architecture/composant peut permettre d'impliquer fournisseurs et clients dans l'apprentissage localisé relatif au développement de composants spécifiques : à cet égard, les auteurs citent Woolsey (1974) à propos du développement du Boeing 777, qui a vu l'implication de *lead customers* dans le développement de composants clefs de cet avion. L'utilisation d'une architecture produit modulaire, peut ainsi fournir un cadre favorable à l'implication de *lead users* dans le développement produit (von Hippel, 1988).

## 1.4 La question de l'émergence des *design rules* :

L'objet de la modularité est principalement de réduire la consommation de ressources et le temps de conception et de développement. La faible interdépendance entre tâches est ici censée faciliter le développement produit (von Hippel, 1990). Mais des ressources sont aussi nécessaires pour intégrer les modules et composants dans un produit complexe. Certains avancent que la firme doit savoir plus que ce qu'elle produit, car les systèmes complexes réclament une connaissance supplémentaire pour réaliser l'intégration (Brusoni et Prencipe, 2001). D'autres avancent que la capacité intégrative est nécessaire pour éviter d'être piégé au cas où les innovations conduiraient à des changements architecturaux. Baldwin et Clark (2000) définissent six opérateurs modulaires, qui constituent le répertoire de base des actions qui peuvent être conduites dans le cadre de n'importe quelle conception modulaire :

- *splitting* ou éclatement d'un système en plusieurs modules,
- *substituting* ou remplacement d'un concept de module par un autre,
- *augmenting* ou augmentation – ajout d'un nouveau module à un système,
- *excluding* ou exclusion d'un module du système,
- *inverting* ou inversion pour créer de nouvelles règles de conception ou *design rules*,
- *porting* ou portage d'un module vers un autre système.

Pour Baldwin et Clark (2000), les concepts modulaires ne peuvent toutefois pas être complètement définis lorsqu'ils émergent de l'esprit des architectes : « Quand un concept est modularisé pour la première fois, les connaissances des concepteurs concernant les interdépendances sont habituellement imparfaites, et en conséquence les règles de conception seront incomplètes. Les règles de conception incomplètes font apparaître des interdépendances imprévues, qui à leur tour nécessiteront des consultations et des itérations entre les concepteurs des modules cachés et les architectes du système. L'intégration et le test de ces concepts seront conduits avec difficulté et les concepts eux-mêmes risquent d'être des échecs. Mais à mesure que les propriétés du système et les modules sont mieux comprises, les règles de conception deviendront plus complètes » (Baldwin et Clark, 2000 ; pp. 76 et 77).

Nous remarquons que le processus d'émergence de ces *design rules*, a toutefois été très peu étudié par la littérature. De plus, et à l'exception des réserves exprimées par Baldwin et Clark (2000) que nous venons d'évoquer, les auteurs considèrent généralement ces règles de conception comme

étant, pour une large part, clairement définies dès le début du processus de conception. Par ailleurs, la littérature sur la modularité traite presque exclusivement du processus de modularisation d'une architecture produit comme d'un processus de dé-intégration d'architecture, ainsi que le décrivent Baldwin et Clark (2000). Prenant l'exemple des ordinateurs, ces auteurs expliquent qu'un module est au départ un élément de la structure (plus exactement un nœud ou « *knot* ») à l'intérieur d'un concept de système interdépendant. A ce stade, il est couplé au système par de nombreuses interdépendances, mais les concepteurs peuvent le « voir » comme un « organe » séparé à l'intérieur du « corps » du système. C'est à ce moment-là que le composant se voit attribuer un nom, en lien avec son rôle dans le système. Une fois le module nommé et son rôle défini, le composant devient un objet d'étude. Une connaissance spécifique relative à celui-ci s'accumule progressivement, alors qu'il reste interdépendant et connecté au reste du système. Des concepteurs spécialisés dans ce composant apparaissent, et commencent à être recherchés par les équipes de conception (les exemples de composants ayant suivi cette trajectoire, sont nombreux : microprocesseurs, systèmes d'exploitation, mémoire vive (RAM), lecteurs de disque, interfaces graphiques, applications Web, etc...). A ce stade, Baldwin et Clark (2000) considèrent que le composant est dans un état « pré-modulaire », état dans lequel il peut rester longtemps, mais un architecte peut décider de construire un concept modulaire et une structure de tâche dans laquelle le composant est un module. Cela conduit à remplacer les liens de coordination entre modules cachés (*hidden modules*) par des règles de conception, une intégration ex-post et une phase de test. Ensuite, les concepteurs de chaque module peuvent travailler indépendamment des concepteurs des autres modules, sans besoin de coordination quotidien. Les développeurs de module vont alors travailler en partie à développer des tests au niveau du module seul, ce qui est indispensable dans le cadre d'une relation contractuelle, pour déterminer dans quelle mesure le travail demandé a été réalisé par l'équipe concernée. Enfin, quand les performances du module sont bien connues et analysables, il peut être commercialisé.

## **1.5. Deux exemples emblématiques dans l'industrie informatique : les ordinateurs IBM s/360 et le système d'exploitation Unix.**

### *1.5.1 Les ordinateurs IBM s/360 :*

Beaucoup d'industries ont connu un certain degré de modularité dans leur processus de production, et un nombre croissant d'entre elles tentent d'étendre la modularité à l'étape de conception. Pour cela, elles cherchent à apprendre de l'industrie informatique. En informatique,

les capacités de stockage et les vitesses de traitement ont explosé, alors que les coûts sont restés stables ou ont baissé, améliorations qui ont résulté d'une complexité croissante du produit : l'ordinateur moderne est constitué d'un nombre élevé d'éléments travaillant de concert. La modularité a permis aux entreprises de gérer cette technologie de plus en plus complexe. En découpant un produit en sous-systèmes, ou modules, les concepteurs, fournisseurs et autres utilisateurs ont gagné énormément de flexibilité (Baldwin et Clark, 2000). Différentes entreprises peuvent prendre la responsabilité de modules séparés, et être certaines que le produit qui émergera de leur effort collectif sera fiable.

Le premier ordinateur modulaire le System/360, lancé par IBM en 1964, en est la parfaite illustration : les modèles précédents possédaient leurs propres systèmes d'exploitation, processeurs, périphériques et logiciels d'application. A chaque nouveau modèle, un constructeur informatique, devait, pour tirer profit des dernières évolutions technologiques, développer des composants et des logiciels spécifiques, tout en maintenant les anciens modèles. Les utilisateurs qui changeaient de machine, devaient réécrire tous leurs programmes, avec le risque de perte de données dû à une mauvaise récupération par les nouveaux programmes, ce qui limitait leur désir de changer de machine. Les développeurs de la gamme 360, conçurent une gamme d'ordinateurs de taille variable pouvant convenir à différentes applications, utilisant les mêmes ensembles d'instruction et pouvant partager les mêmes périphériques. Pour atteindre cette modularité, ils appliquèrent les principes de la modularité en conception : les concepteurs divisèrent les concepts des processeurs et des périphériques en information visible et cachée (cf. définitions). IBM mis sur pied un Central Processor Control Office, qui établit et renforça les règles de conception visibles qui déterminaient comment les différents modules de la machine devaient travailler ensemble. Les équipes de conception réparties dans le monde entier durent se plier scrupuleusement à ces règles. Mais chaque équipe avait un contrôle complet sur les éléments cachés de conception dans son module.

L'emploi de cette approche par IBM aboutit à un succès colossal, beaucoup de ses concurrents étant forcés de quitter le marché, ou de se réfugier sur des niches. Mais sur le long terme, la modularisation finit par miner aussi la prédominance de IBM, de nouvelles entreprises produisant progressivement des modules - imprimantes, terminaux, mémoire, logiciel, et finalement les unités de traitement central elles-mêmes - compatibles avec les machines IBM. En suivant les règles d'IBM, et en se spécialisant sur un seul module, chaque entreprise ou unité pouvait approfondir ses travaux. Le fait d'avoir de nombreuses entreprises focalisées sur un seul module,

permet de nombreuses expérimentations parallèles. Les concepteurs de module étaient libres d'essayer toute une gamme d'approches tant qu'elles respectaient les règles de conception. Pour une telle industrie, dans laquelle l'incertitude technologique est élevée, et la meilleure façon de procéder souvent inconnue, plus il y a d'expérimentations, et plus les concepteurs ont de flexibilité pour développer et tester les modules expérimentaux, plus l'industrie sera capable d'aboutir rapidement à des versions améliorées. Cette liberté d'expérimenter un concept de produit est ce qui distingue les fournisseurs modulaires des sous-traitants ordinaires. Ainsi, une équipe de concepteurs de lecteurs de disques doit obéir aux spécifications générales d'un PC, telles que les protocoles de transmission de données, les spécifications de taille et de forme du matériel, en respectant les standards relatifs aux interfaces, pour que le module fonctionne dans le système. Mais, d'un autre côté, les membres de l'équipe peuvent concevoir le lecteur de disque de la manière qu'ils pensent être la meilleure. Leurs décisions n'ont pas à être communiquées aux concepteurs des autres modules ou même aux architectes du système, créateurs des règles visibles. Des fournisseurs de lecteurs de disque peuvent ainsi expérimenter leur version du module avec des approches ingénieriques complètement différentes.

### *1.5.2 Le système d'exploitation Unix :*

L'exemple du système d'exploitation Unix, nous intéresse à plus d'un titre, tout d'abord parce qu'il concerne l'univers du logiciel, ensuite parce qu'il illustre la façon avec laquelle la diffusion d'un produit doté d'une architecture modulaire peut impacter la structure d'une industrie, et enfin parce qu'il montre comment le caractère modulaire de ce logiciel a permis à un grand nombre d'acteurs de participer au processus d'innovation, parmi lesquels un certain nombre d'utilisateurs.

Pour Baldwin et Clark (2000), la force d'Unix est d'être modulaire à la fois dans sa propre structure de conception (modularité statique), et dans la façon avec laquelle il gère les ressources de l'ordinateur (modularité dynamique). La modularité « statique » est en fait la modularité traditionnelle : les composants individuels d'Unix ont été partitionnés avec précaution afin que les tâches de conception sur chaque module puissent être réalisées indépendamment. Ainsi, Unix était facile à mettre à jour et à « augmenter », au sens où nous l'avons défini quand nous avons décrit les six opérateurs de la modularité. Sa conception invitait au « bricolage », et ses utilisateurs ont rapidement développé des mécanismes de partage de solutions aux problèmes rencontrés le plus fréquemment. Les utilisateurs purent améliorer des modules, ou même en ajouter. La modularité « statique » rendait aussi le « portage » d'Unix plus facile (le portage informatique

consiste à mettre en œuvre un logiciel, une fonctionnalité, voire un système d'exploitation comme Unix, dans un autre environnement matériel ou logiciel que celui d'origine). Mais ce qui rendait Unix unique à l'origine, c'était sa façon de gérer les ressources de l'ordinateur, en particulier la modularité de sa gestion dynamique de la mémoire. La gestion dynamique de la mémoire affecte grandement la vitesse d'un ordinateur. Auparavant, cette gestion était souvent encadrée dans le code des programmes d'application, alors que cette fonction dépendante du logiciel devait être réalisée par tous les programmes fonctionnant sur un certain type de machine. A ce titre, cette fonction était un candidat évident à l'« inversion ». Unix créa alors une gestion de la mémoire modulaire, au travers de la définition de *images* et de *reentrant processes* et en utilisant un interpréteur de commandes, appelé le *Shell*. Unix s'est finalement imposée ainsi que sa philosophie, qui était une affirmation des principes de modularité : découper le code en unités fonctionnelles aussi petites que possible et trouver la meilleure mise en œuvre de cette fonction, masquer les contenus d'un module en présentant ses inputs et ses outputs sous forme de texte (« l'interface universelle »), et créer de grands systèmes à partir de modules élémentaires.

## **1.6. Les conséquences de la modularité sur les organisations et la structure d'une industrie :**

Maintenant que nous avons défini et illustré la notion de modularité, nous évoquons ci-après un certain nombre de travaux qui s'intéressent aux implications organisationnelles de la modularité.

### *1.6.1 Réseaux de fournisseurs de systèmes modulaires :*

La spécialisation verticale encouragée par les systèmes modulaires conduit à l'établissement de réseaux de fournisseurs. Ces réseaux sont appelés *modular cluster* (Baldwin et Clark, 2000). Fondamentalement, pour Langlois et Roberston (1992) deux types de réseau sont possibles : le premier est centralisé autour d'une firme leader, comme dans l'automobile japonaise. Le deuxième est décentralisé : il y a dans ce cas, des clients qui assemblent selon leurs besoins, des fournisseurs de chaque type de composant, et des fournisseurs de sous-assemblages utilisés par l'un des fournisseurs de composant. A la différence des réseaux centralisés, dans lesquels les standards de compatibilité sont édictés par la firme leader qui joue le rôle d'architecte unique, dans les réseaux décentralisés les standards sont déterminés conjointement par les producteurs de composants et les utilisateurs/assembleurs au travers de négociation ou de processus de marché. Aucun membre ne contrôle seul le réseau : des variations de composant peuvent même échouer

si les utilisateurs et d'autres fournisseurs sont verrouillés par des standards existants, car les coûts de transfert sont plus élevés que les bénéfices escomptés. Quand les fournisseurs de composant (spécialement logiciel) ne peuvent pas offrir aux clients assez de variété dans un système modulaire, les firmes qui rencontrent le plus de succès sont celles qui abandonnent une stratégie propriétaire en faveur de l'appartenance à un réseau de concurrents employant un standard commun de compatibilité (Langlois et Robertson, 1992).

A cela s'ajoute, dans le cas du logiciel, le phénomène du leadership de plate-forme ou *platform leadership* (Gawer et Cusumano, 2002, p6), comme nous l'avons expliqué dans le chapitre I, « L'industrie du logiciel ». Ces auteurs affirment que, dans les secteurs où la modularité est la règle, et spécialement l'industrie du logiciel, de plus en plus d'entreprises veulent être des leaders de plate-forme. Cela signifie que ces entreprises adoptent des « stratégies de plate-forme », afin que leurs produits constituent les fondations sur lesquelles d'autres entreprises bâtiront leurs propres produits ou services appelés alors « produits, ou services, complémentaires ». On rencontre des leaders de plate-forme en informatique que ce soit dans le domaine du hard, avec par exemple Intel, ou dans celui du logiciel, avec entre autres Microsoft et son système d'exploitation Windows. Les leaders de plate-forme bénéficient des innovations développées par les entreprises qui fournissent des *produits ou services complémentaires*, tout en en étant dépendants. La disponibilité d'un grand nombre de produits complémentaires ajoute de la valeur au *core product*, ou produit cœur, qui constitue la plate-forme à laquelle les produits complémentaires viennent se rattacher. Pour Microsoft, l'existence de très nombreux logiciels « complémentaires » compatibles Windows, encourage ainsi les clients potentiels à acheter des ordinateurs équipés de son système d'exploitation. Dans l'univers du logiciel (cf. chapitre 1, « L'industrie du logiciel »), pour chaque nouveau type d'application, quelques leaders émergent, après que le marché a atteint une certaine maturité. Ces leaders du marché sont généralement aussi des leaders de plate-forme, comme dans le cas des systèmes d'exploitation avec, par exemple, Microsoft, ou dans celui des progiciels intégrés avec par exemple, SAP. Gawer et Cusumano (2002) avancent que les plateformes étant faites de composants qui interagissent entre eux selon des interfaces standard, la guerre des standards fait obligatoirement partie des « stratégies de plate-forme ». Parmi les nombreux exemples de « guerre des standards », Gawer et Cusumano (2002) citent la guerre Macintosh/PC compatible Windows. Dans le logiciel, l'émergence de nouveaux types d'application s'accompagne donc généralement d'une stratégie de leadership de plate-forme, ce qui signifie qu'un des enjeux des éditeurs de logiciel proposant de telles applications est bien de réaliser une

plate-forme logicielle susceptible d'attirer les fournisseurs de logiciels complémentaires, en imposant ses propres standards.

### 1.6.2 *Architecte ou modulier, deux stratégies exclusives ?*

Dans les marchés de produits dotés d'architecture modulaire, les revenus sont beaucoup plus dispersés que dans les industries traditionnelles. Il est en effet très différent de faire partie d'un *modular cluster* de centaines d'entreprises dans une industrie en évolution constante, que d'être une des entreprises dominantes d'une industrie stable. La nature duale de la structure d'un marché modulaire impose aux managers de choisir entre deux stratégies : être architecte, créant l'information visible, ou *design rules*, pour un produit fait de modules. Ou être un concepteur de modules, conformes à l'architecture, aux interfaces, aux protocoles de test des autres. Les deux stratégies diffèrent sur certains points :

1. Pour un architecte, les avantages concurrentiels viennent de la capacité à attirer les concepteurs de modules vers ses *design rules*, en les convainquant que cette architecture prévaudra sur le marché.
2. Pour le fournisseur de module, l'avantage concurrentiel vient de la maîtrise de l'information cachée du concept de module en question et de la capacité à opérer une exécution supérieure. A mesure que les opportunités émergent, le fournisseur de module doit évoluer rapidement pour satisfaire un besoin, et ensuite atteindre de nouveaux niveaux de performance, à mesure que le marché devient plus dense, ou changer de marché.

On observe toutefois que les « architectes » sont aussi, souvent, les fournisseurs de certains modules, comme par exemple Intel et Microsoft. L'exemple donné par ces entreprises incite à penser que les entreprises devraient chercher à contrôler les règles visibles de conception en développant des architectures propriétaires et laisser les détails des modules cachés aux autres. Il est vrai que la position d'architecte est puissante, et peut être très profitable. Mais un challenger peut s'appuyer sur la modularité pour combiner ses propres capacités avec celles des autres et supplanter un architecte. C'est ce qui s'est produit sur le marché des stations de travail dans les années 1980 (Baldwin et Clark, 2000).

### 1.6.3 Correspondance entre architecture produit et structure organisationnelle :

La question du lien entre modularité des architectures produit et modularité de l'organisation en charge de la conception et de la production de ce produit, est un thème récurrent dans la littérature. On distingue à ce sujet deux visions opposées, l'une selon laquelle à une structure produit correspond naturellement et idéalement une structure organisationnelle elle-même modulaire, i.e. constituée de modules organisationnels faiblement couplés en charge, chacun et indépendamment des autres, de concevoir et développer un des modules constitutifs du produit. Selon l'autre vision, l'isomorphisme entre architecture produit et architecture organisationnelle n'est pas observé systématiquement dans les cas concrets d'architectures produit modulaires, et ne constitue pas un impératif de succès d'une stratégie produit modulaire (Fixson, 2003).

La première vision est partagée par des auteurs tels qu'Orton et Weick (1990), pour qui une architecture produit modulaire fournit une forme de coordination encadrée, qui rend possible le développement autonome et concourant de composants, par des organisations faiblement couplées. Pour Baldwin et Clark (1997), la modularité augmente à la fois le taux d'innovation, et le degré d'incertitude dans le processus de conception. Pour se préparer aux changements drastiques du marché, il faut donc que les managers soient capables de choisir parmi un ensemble souvent complexe d'options financières ou liées à la technologie, ou encore liées aux compétences. En conséquence, pour Baldwin et Clark (1997), les managers qui évoluent dans un système modulaire, doivent aussi reconcevoir leur organisation interne : pour créer des modules supérieurs, ils ont besoin de pouvoir bouger rapidement sur le marché, et de faire usage des technologies en évolution rapide, tout en s'assurant de la compatibilité avec l'architecture définie. Pour Baldwin et Clark (1997), comme pour Sanchez et Mahoney (1996), la réponse à ce dilemme, est la modularité dans l'organisation. Sanchez et Mahoney (1996) analysent la capacité des interfaces standardisées entre composants d'un produit, à encadrer la coordination des processus de développement de produit. La « coordination encadrée » ou *embedded coordination* créerait selon eux une « coordination horizontale » sans qu'il y ait besoin d'exercer une autorité – permettant une coordination efficace des processus sans le couplage fort de structures organisationnelles.

## 1.7. Modularité et innovation :

### 1.7.1 Introduction :

La conception selon l'ingénierie traditionnelle essaie d'obtenir le plus haut niveau de performance produit en prenant en compte un certain nombre de contraintes de prix et de performance : cela conduit à des concepts de composants fortement couplés et intégrés. En conséquence, les processus correspondants demandent une forte coordination managériale. Les firmes qui conçoivent de manière modulaire, utilisent des éléments faiblement couplés en spécifiant des interfaces de composants standardisées qui définissent les relations fonctionnelles, spatiales, et autres, entre composants (Sanchez et Mahoney, 1996). Une fois spécifiées, elles ne peuvent changer pendant une période donnée du processus de développement (les interfaces sont définies, puis figées pendant un certain temps). Cette période peut se limiter à des étapes clés du développement d'une nouvelle architecture produit (Cusumano et Selby, 1995), ou couvrir l'ensemble du cycle de vie commerciale d'un produit (Sanchez, 1995). Pendant cette période, les processus de développement se coordonnent naturellement à partir du moment où les spécifications d'interface de composant standardisé sont respectées, sans recourir à l'autorité managériale, ce qui conduit Sanchez et Mahoney (1996) à parler, nous l'avons vu, de « coordination encadrée ». Une architecture produit modulaire (Sanchez 1994 ; Ulrich et Eppinger, 1995) est cependant flexible car des variations de produits peuvent être démultipliées, notamment, par substitution (Garud et Kumaraswamy, 1995). Le faible couplage permet le « *mixing and matching* » de composants qui peut aboutir à des variations de performances, fonctions, ou fonctionnalités (Sanderson et Uzumeri, 1990). Ainsi les architectures produit modulaires peuvent être une source de flexibilité stratégique (Sanchez, 1995), quand elles permettent de répondre plus rapidement aux changements des marchés et des technologies, grâce à de nouvelles combinaisons de composants. La modularisation et le « *mixing and matching* » qui en découlent, apparaissent dans de nombreux secteurs tels que l'aérospatiale, l'automobile, l'électronique grand public, les appareils domestiques, la microinformatique, et le logiciel (Parnas, Clements et Weiss, 1985 ; Cusumano, 1991 ; von Hippel, 1994). Toutefois, il convient de souligner que dans la construction automobile, par exemple, le produit n'est que partiellement modulaire, car s'il existe des sous-ensembles localisés pour lesquels la modularité peut être mise en œuvre avec succès, on ne peut pas réellement parler de modularité de conception dans ce secteur, contrairement à l'informatique (Fixson, 2003).

### 1.7.2 L'innovation relative aux produits modulaires :

Un système modulaire est ouvert à l'innovation d'un certain type. Ainsi, un réseau décentralisé peut permettre l'essai de nombreuses approches alternatives simultanées, conduisant à un apprentissage essai-erreur rapide. Ce genre d'innovation est très important dans le cas où la technologie évolue rapidement et où il y a un degré élevé d'incertitude concernant le marché et/ou la technologie (Nelson et Winter, 1977). Dans un réseau décentralisé, il y a beaucoup plus de points d'entrée pour de nouvelles firmes, donc pour de nouvelles idées, que dans une industrie intégrée verticalement produisant des appareils fonctionnellement similaires. De la sorte, un système modulaire peut progresser plus vite technologiquement, spécialement pendant les périodes d'incertitude. Une autre raison, pour laquelle l'innovation peut être favorisée par la modularité réside dans la division du travail. Un réseau avec un standard de compatibilité promeut l'innovation autonome, i.e. l'innovation demandant peu de coordination à chaque étape. En permettant, aux producteurs spécialistes (et parfois aux utilisateurs spécialistes) de concentrer leur attention sur des composants particuliers, un système modulaire met ainsi la division du travail au service de l'innovation. On s'attend à ce que l'innovation procède de la manière suggérée par Hughes (1983) : avec des composants « goulets d'étranglement », ceux qui conduisent à une satisfaction client accrue, comme points focaux de changement. De même, pour Baldwin et Clark (2000) dans un système modulaire, constitué d'acteurs indépendants, les options de conception ne sont plus contrôlées par un ou quelques acteurs, mais sont largement dispersées au travers du *cluster* modulaire de firmes indépendantes, ce qui aboutit à la création de nombreux domaines, où des individus ou des firmes peuvent répondre à des incitations locales de capture ou de création de valeur. Le développement de concept centralisé laisse alors la place à ce que les auteurs appellent la « *design evolution* ». La trajectoire du système dans son ensemble devient alors une agrégation des trajectoires individuelles de chaque module.

A partir de l'étude des cas de la stéréophonie et de la microinformatique, Baldwin et Clark (2000) constatent qu'à chaque fois, l'industrie a adopté une structure modulaire avec un standard commun de compatibilité plutôt qu'une structure de produits « packagés » concurrents. Dans les deux cas, les grandes entreprises ont essayé de s'approprier les rentes liées à l'innovation par une approche propriétaire. Mais elles ont échoué et les entreprises s'appuyant sur un réseau externe de concurrents et de fournisseurs ont rencontré beaucoup plus de succès. Ainsi, Columbia a encouragé la production de disques 33 tours et de lecteurs de disques, et IBM permis à Microsoft d'accorder des licences de MS-DOS, sur une grande échelle. Ces firmes devinrent des acteurs

significatifs dans des réseaux qu'elles ne contrôlaient pas, et en ont retiré un profit supérieur à ce qu'il aurait été si elles avaient essayé de vendre un produit propriétaire. Teece (1986) a suggéré des pistes selon lesquelles la volonté de s'approprier ces rentes peut conduire à l'intégration verticale. Les cas de la stéréophonie et de la microinformatique suggèrent la possibilité inverse : la même volonté peut conduire à la désintégration verticale (et horizontale). Dans les deux cas, les fans et les passionnés, ayant des goûts plus sophistiqués et un plus grand désir de payer, ont joué un rôle important. Ces passionnés ont testé les limites des systèmes, et aidé à identifier les goulets d'étranglement qui sont devenus les points focaux de l'innovation. Dans de nombreux cas, ces individus se sont lancés sur le marché, pour fournir (et améliorer) ces composants. Dans les deux cas, la compatibilité des matériels et des logiciels constituait un des problèmes majeurs. En dehors de toute externalité de réseau, la modularité des systèmes microinformatique et stéréo, a permis aux producteurs de participer à un système qui était capable de couvrir l'espace produit – et ainsi, de générer une plus grande demande des consommateurs – qu'un système d'entités « prépackagées » concurrentes. Dans les premières étapes de développement d'un produit, l'expérimentation est un problème beaucoup plus important que la coordination. Et l'apprentissage rapide par essai-erreur est une force d'un réseau décentralisé (Baldwin et Clark, 2000), ce qui constitue donc une supériorité des architectures produit modulaires sur les architectures produit intégrées.

Ainsi, un processus de conception de produit modulaire peut, nous l'avons évoqué, permettre d'améliorer sa connaissance du marché en permettant de proposer des variations d'un produit plus vite et à un coût plus faible. Cela peut en effet faciliter un processus évolutionnaire de *market research* en temps réel (Sanchez et Sudharshan ; 1993) qui aide à la création accélérée de connaissances relatives au marché dans l'entreprise (Baldwin et Clark, 1994). Le découplage des processus d'apprentissage architecture/composant, peut aussi créer un environnement efficace permettant d'impliquer fournisseurs et clients dans l'« apprentissage localisé », pour le développement de composants spécifiques. L'utilisation, déjà évoquée plus haut, par Boeing d'un processus de conception modulaire dans le développement de son modèle 777 (Woolsey, 1994), a ainsi créé un environnement qui facilita l'implication de « *lead customers* » pour le développement de concepts améliorés de composants clefs de l'avion, affectant directement leur usage par les clients. L'utilisation d'architecture produit modulaires peut ainsi fournir un cadre favorable à l'implication de *lead users* (von Hippel, 1988) dans le développement de produit.

En ce qui concerne, l'apprentissage au niveau de l'architecture du produit, les avis divergent, et l'opinion de Sanchez et Mahoney (1996) semble minoritaire parmi les chercheurs qui ont travaillé sur la modularité. Sanchez et Mahoney (1996) estiment en effet que le découplage de l'apprentissage architectural de l'apprentissage au niveau des composants peut permettre aux organisations concernées, de se dégager des demandes court-terme et de l'apprentissage « myope » (« *myopic learning* ») que décrivent Levinthal et March (1993). A l'inverse, de nombreux auteurs adoptent le point de vue développé par Henderson et Clark (1990) selon lequel la focalisation au niveau de chaque composant peut freiner l'innovation architecturale. On retrouve ce point de vue sous une forme un peu différente chez Langlois et Robertson (1992), qui expliquent que l'innovation systémique est plus difficile dans un système modulaire, et même parfois, non désirable en ce qu'elle détruirait la compatibilité entre composants.

En conclusion sur ce point, nous constatons que le choix de la modularité conduit les entreprises à mettre l'accent sur l'apprentissage architectural, et les conduit généralement à se focaliser sur un ou quelques composants clefs critiques quant à la performance du système, et à laisser la conception et le développement des autres composants, à des entreprises tierces. L'innovation au niveau des composants (modules) est facilitée par une architecture produit modulaire, qui facilite l'apprentissage localisé et l'implication des clients et des fournisseurs, tandis que les chercheurs ne sont pas d'accord sur le point de savoir si l'innovation au niveau architectural est, quant à elle, favorisée ou handicapée par ce type d'architecture.

### 1.7.3 Architecture produit et exploration des marchés :

Sanchez et Sudharashan (1993) décrivent une stratégie qu'ils appellent « *real-time market research* » et qui, selon eux, peut être déployée très efficacement dans le cas d'une architecture produit modulaire. Selon ce schéma, la firme introduit un produit, évalue la réponse du marché, puis développe et lance un produit amélioré de manière incrémentale, extrêmement rapidement. Les bénéfices d'une architecture modulaire pour explorer un marché et le « *fine tuning* » d'un produit sont aussi décrits par Langlois et Robertson (1992). La modularité favorise ainsi les découvertes et les changements dans la phase d'utilisation grâce au *mixing and matching* d'éléments prédéterminés (la stéréo par Langlois et Robertson en 1992), aux mises à jour, aux *add-ons*, et aux adaptations (110 ou 220 volts). Il s'agit là de ce que la littérature appelle la « modularité d'usage » (Baldwin et Clark, 2000). Langlois et Robertson (1992) ont abordé la question de l'innovation dans les systèmes modulaires, en étudiant les industries des composants microinformatiques et

stéréophoniques. Pour eux, si on a beaucoup étudié les effets des conditions dans lesquelles l'offre est produite sur la structuration de la nature d'une industrie, cela a été beaucoup moins étudié dans le cas de la demande. Pour les auteurs, la structure d'une industrie et les caractéristiques de ses firmes peuvent varier grandement selon la façon avec laquelle les consommateurs définissent ce produit : avec le temps, la nature de ce que les clients croient être l'essence d'un produit, change souvent. Les clients peuvent ajouter certains attributs, et en rejeter d'autres, ils peuvent aussi combiner le produit avec un autre produit précédemment considéré comme différent. A l'inverse, un produit que des clients avaient traité comme une entité, peut être divisé en groupe de sous-produits qu'ils peuvent arranger en combinaisons variées selon leurs préférences. La nature d'une industrie et l'étendue de l'intégration verticale ne dépendraient donc pas uniquement de la minimisation des coûts de transaction et de production, mais aussi de ce que les clients souhaitent (Langlois et Robertson, 1992).

#### *1.7.4 Limites de l'architecture modulaire :*

Les architectures modulaires permettent l'optimisation des caractéristiques des performances locales : d'abord, parce qu'une architecture modulaire peut permettre l'usage de composants standard (qui bénéficie du fait que le fournisseur a intégré toute l'expertise et l'expérience de l'utilisation passée) ; ensuite, quand un composant standard n'est pas disponible et doit être complètement développé, une architecture modulaire permet au composant, d'être conçu, testé, amélioré, d'une manière focalisée. Ces bénéfices conduisent à des performances de composant plus élevées. Par contre dans certains cas, Ulrich (1995) estime que l'optimisation des performances globales peut conduire à sacrifier la modularité de l'architecture. Ainsi, l'architecture intégrale est utile, par exemple, pour partager des fonctions entre composants. Cela explique pourquoi on emploie parfois des architectures intégrales, pour atteindre des coûts unitaires très bas.

#### *1.7.5 Architecture et processus de conception :*

Ulrich (1995) examine l'impact des différentes architectures produit lors des phases successives du processus de conception : développement du concept, conception au niveau du système (développement de l'architecture produit, et attribution des tâches à l'équipe de développement produit), conception détaillée (planning du processus de conception de composants, de test et de production), test du produit et raffinement (assemblage et test des prototypes, et mise en œuvre

de tout changement dans les concepts de composant). Selon l'auteur, l'architecture produit a l'impact suivant sur ces différentes phases :

### **1. Conception du système :**

Pour une architecture modulaire, cette phase nécessite un investissement fort. Il s'agit de définir avec précaution les interfaces des composants, en spécifiant les standards et protocoles associés. Les performances et les critères d'acceptation sont définis pour chaque composant, correspondant à l'élément fonctionnel implémenté par ce composant. La conception du composant est fréquemment attribuée aux spécialistes internes ou externes. Le leader de l'équipe de développement peut être considéré comme un « architecte système poids lourd » ou *heavyweight project manager* (Wheelwright et Clark, 1992). Pour une architecture intégrée, cette étape absorbe peu d'effort. Il faut établir des objectifs clairs concernant la performance du système complet, et la division du système en petit nombre de sous-systèmes intégrés, généralement attribués à des équipes multidisciplinaires. Le leader de ces équipes peut être considéré comme un « intégrateur système poids lourd ».

### **2. Conception détaillée :**

Pour une architecture modulaire, il faut surveiller les progrès de chaque composant, en fonction des objectifs fixés. Les équipes de développement sont considérées comme des fournisseurs, car l'interaction est structurée et peu fréquente. Le test de chaque composant peut être effectué indépendamment, et des objectifs clairs indiquent la complétion de la tâche de conception de chaque composant. Pour une architecture intégrée, les concepteurs de composant forment ensemble une « *core team* », et interagissent continuellement, afin d'analyser la performance de leur sous-système, et d'effectuer les changements rendus nécessaires par le couplage d'interface. Les composants atteignent leurs objectifs, selon leur interaction et non en fonction de critères pré-spécifiés. Le test de chaque composant ne peut s'effectuer isolément. Les sous-systèmes doivent être assemblés et testés comme un tout.

### **3. Test produit et raffinement :**

Pour l'architecture modulaire, c'est une activité de contrôle. Les tests visent à détecter des interactions inattendues entre composants. Ces interactions sont vues comme des bugs, et leur résolution est généralement réalisée par la modification d'un ou deux composants. Pour

l'architecture intégrée, c'est une activité de réglage (*tuning*). Les éventuels changements touchent de nombreux composants, et cette phase prend plus de temps que dans le cas de l'architecture modulaire (Ulrich, 1995).

#### 1.7.6 Modularité et innovation : conclusion.

La littérature (Ulrich, 1995 ; Baldwin et Clark, 1997 et 2000 ; Fixson, 2003) nous enseigne que les architectures modulaires semblent mieux adaptées aux environnements incertains, tant au niveau du marché qu'au niveau technologique, en ce qu'elles permettent des développements plus rapides, car décentralisés et concourants, et en ce qu'elles permettent des apprentissages localisés. Elles facilitent dans certains cas, l'implication des fournisseurs et des utilisateurs au processus d'innovation, ainsi que l'exploration de nouveaux marchés. A l'inverse, elles requièrent, de la part de l'architecte de l'innovation, des investissements lourds dans la phase de conception architecturale, pour aboutir à un niveau suffisant de définition des *design rules*. Elles requièrent aussi de sa part, toujours selon la littérature, la capacité à identifier, puis à contractualiser et enfin à collaborer avec des organisations faiblement couplées. Enfin, une fois les *design rules* établies, la question de savoir dans quelle mesure l'innovation architecturale est favorisée ou handicapée par une architecture produit modulaire n'est pas tranchée par la littérature, tant les avis divergent sur ce point.

### 1.8 Conclusion et questions soulevées par la littérature sur la modularité :

L'analyse de la littérature nous conduit à nous interroger : tout d'abord la tendance croissante à la *modularisation* des technologies et à la désintégration des systèmes implique la production de nouveaux types de connaissances (Steinmueller, 2002), telles que normes, standards, savoirs d'intégration (Shapiro et Varian, 1999). La modularité des architectures produit passe ainsi par la définition de *design rules* ou règles de conception, qui, comme le mentionnent Baldwin et Clark (2000), ne sont pas complètement et parfaitement définies lors de leur première élaboration par leurs concepteurs. La définition précise des *design rules* est donc le résultat d'un processus, processus qui a été fort peu traité dans la littérature. Ces nouveaux types de connaissance sont nécessaires à l'intégration d'ensembles faiblement couplés. Comment ces connaissances émergent-elles ? Qui les construit ? Comment sont élaborées les *design rules*? Ces questions restent ouvertes.

La trajectoire conduisant à la modularisation a, quant à elle, été étudiée sous l'angle de la désintégration ou « dé-intégration » (Baldwin et Clark, 2000), généralement progressive d'une architecture intégrée. Or, le secteur du logiciel présente un certain nombre de particularités qui peuvent conduire à considérer un autre processus de conception d'une architecture produit modulaire : il s'agit du phénomène qui consiste à concevoir une innovation en associant différents modules, déjà développés, mais indépendants les uns des autres, et nécessitant donc un travail d'adaptation et d'intégration. Ce phénomène n'est pas celui de modularité d'usage observé dans le secteur de la stéréophonie ou de la microinformatique tels que nous les avons décrits précédemment : la modularité d'usage signifie que l'utilisateur, entreprise ou individu, choisit un certain nombre de modules et les associe (« *mix and match* ») pour bâtir son propre produit. Elle est décrite par la littérature comme une opération se déroulant dans des conditions où les *design rules* ont été clairement édictées, et où les interfaces entre modules sont parfaitement définies, tout comme les modules dont les utilisateurs peuvent évaluer a priori les performances en tant qu'élément séparé, et en tant que constituant du système modulaire qu'il cherche à bâtir (comme par exemple, les composants d'un micro-ordinateur à partir de la fin des années 1980). A contrario, nous nous intéressons à des composants, des logiciels innovants, qui ont été conçus indépendamment les uns des autres, par des équipes de recherche universitaires ou de grandes entreprises, équipes qui ont essaimé par la suite comme c'est souvent le cas dans l'industrie du logiciel.

La création de solutions modulaires résulterait alors de l'agrégation de logiciels qui peuvent théoriquement s'interfacer entre eux, s'ils sont conçus pour un même environnement (comme l'environnement créé par la combinaison des microprocesseurs Intel et du système d'exploitation Windows de Microsoft), et, a fortiori, s'ils emploient des formats de données standard tel que le format XML, par exemple. Mais une phase d'intégration, est nécessaire pour développer de passerelles logicielles, la mise en cohérence des types et formats de données utilisés par chaque brique logicielle, la création d'une application qui fédère ces briques, mais aussi dans le cas des logiciels destinés aux entreprises, la personnalisation de la solution, c'est-à-dire son adaptation aux besoins particuliers du client, à son « métier ». Cette personnalisation des logiciels est pratiquement systématique dans le cas des logiciels de type Progiciel de Gestion Intégré (PGI ou ERP), ou encore de gestion de la relation client (GRC ou CRM). Le processus de conception d'un progiciel innovant doté d'une architecture modulaire, où chaque module a été conçu antérieurement au produit lui-même par des organisations indépendantes, correspond à un processus que nous considérons comme relativement typique de l'industrie logicielle, où

l'innovation est portée par de petites structures très spécialisées (Cusumano, 2004). Ce processus de modularisation a cependant été peu étudié dans la littérature, il y a donc là un champ à explorer.

Dès lors les questions sont nombreuses, qui ont trait à ce processus de modularisation d'un type particulier. Il se pose d'abord la question de l'architecte de la solution : à la différence des cas où le point de départ est constitué par un produit doté d'une architecture intégrée, le rôle de l'architecte n'est pas attribué a priori, qui joue ce rôle ? Un des éditeurs de logiciel, une société tierce (par exemple une société de services informatiques), voire un client, plusieurs acteurs à la fois ? Nous avons vu que le processus de modularisation traditionnel tel qu'il est décrit par la littérature peut aboutir à un système modulaire ou *modular cluster*, où de nombreux acteurs peuvent innover localement, dont parfois les utilisateurs (cf. le cas du logiciel Unix, cité par Baldwin et Clark, 2000 ; ou du Boeing 777 cité par Sanchez et Mahoney, 1996). Quelles compétences doit posséder un tel architecte, notamment dans le cas d'innovations radicales, que nous nous proposons d'étudier ? Comment l'architecture produit évolue-t-elle dans un tel contexte ?

En conclusion, nous allons à chercher à étudier un processus de modularisation original, car peu étudié dans la littérature, mais qui correspond à une situation typique dans l'industrie du logiciel, où les innovations radicales sont conçues par des petites structures organisationnelles très spécialisées, des « start-up technologiques », qui tentent de suivre une stratégie de plate-forme. Nous nous appuyerons sur une étude empirique, ce qui constitue une démarche relativement peu répandue dans le champ de la modularité, comme l'a souligné Fixson (2003).

## **2. LES APPORTS DE LA LITTÉRATURE SUR L'INNOVATION PAR LES UTILISATEURS : LE LOGICIEL COMME CHAMP D'APPLICATION IDEAL ?**

Comme nous l'avons évoqué, le concept de modularité et l'innovation par les utilisateurs constituent deux modèles dominants dans l'informatique en général, et dans le secteur du logiciel en particulier. Ces deux notions sont liées par l'idée que la modularisation d'une architecture logicielle permet d'éclater le processus de conception en un certain nombre d'acteurs, parmi lesquels les utilisateurs. C'est en tout cas le point de vue de la littérature sur l'innovation par les utilisateurs, qui attribue aux utilisateurs un rôle particulier dans le processus de conception et de développement des logiciels. Cela nous conduit à détailler maintenant les apports de cette approche.

Les travaux relatifs à l'innovation par les utilisateurs ont été initiés, puis développés par von Hippel dès le milieu des années 1970. Il a progressivement développé une approche théorique et concrète. Au niveau théorique, il a fait émerger les notions de *lead user* (von Hippel, 1986), et de *stickiness* (ou adhérence) de l'information détenue par les utilisateurs (von Hippel, 1994), et souligné l'hétérogénéité des besoins des utilisateurs à laquelle l'utilisation de *toolkits* (ou boîtes à outils) permet de répondre (von Hippel et Katz, 2002). Fondamentalement, il affirme que certains utilisateurs *lead users* peuvent jouer un rôle clef dans les processus d'innovation, et ce pratiquement quels que soient les secteurs étudiés : high tech (cas des logiciels ou de l'instrumentation médicale), *low tech* (cas de l'entreprise Hilti), ou sports *out-door* (cas du *kite surf*). Ce rôle est renforcé lorsque les utilisateurs peuvent s'appuyer sur une communauté d'utilisateurs active dans le processus (von Hippel, 2005). De plus, von Hippel, Thomke et Sonnack (1999) affirment que l'approche *lead user*, aide les entreprises à surmonter le défi que constitue la génération d'un flux régulier d'innovations radicales, un enjeu économique crucial. Nous nous proposons de présenter ces différents apports, et les questions qu'ils soulèvent.

## 2.1 Les premiers travaux de von Hippel :

Dès 1977, dans l'article "*Has a customer already developed your next product?*" (von Hippel, 1977), von Hippel note que « la stratégie habituelle des entreprises pour développer avec succès de nouveaux produits est de « trouver un besoin et le satisfaire ». Mais qu'est-ce qu'un besoin, exactement, et où pouvez-vous le chercher ? ». L'auteur s'appuie sur une étude de trois ans qui « a permis de faire émerger le fait que l'information relative à un besoin est souvent groupée avec des données relatives à un concept de produit. Ces données ne sont pas toujours accessibles à des professionnels du marketing recherchant seulement des besoins ».

Un des premiers exemples sur lequel von Hippel (1977) fonde son approche est celui des cartes de circuit imprimé (*PC-Boards*) : IBM a conçu la première machine d'insertion de composants de carte de circuit imprimé, dans les années 1950. IBM avait besoin d'une machine pour insérer des composants sur des cartes de circuit imprimé, qui devaient ensuite être intégrées à des ordinateurs. Après avoir conçu et testé le concept en interne, IBM envoya ses plans à un fournisseur de machine local. Ce fournisseur obtint d'IBM le droit de vendre cette machine, ce qui représentait sa première incursion dans l'univers de ce type d'équipement. Plus généralement, von Hippel (1977) recense les différents travaux relatifs aux produits innovants commercialisés par les fabricants de biens d'équipement et d'instruments scientifiques, et en conclut que la

plupart de ceux qui furent commercialisés, avaient été inventés, prototypés et utilisés par des utilisateurs. Von Hippel (1977) note aussi que ce type d'innovation, *user innovation*, que l'on peut traduire par «innovation par les utilisateurs» ou simplement par «innovation utilisateur», joue un grand rôle dans d'autres domaines comme les logiciels informatiques et les matériels utilisés en médecine.

Dans "*A Customer-active paradigm for industrial product idea generation*" (1977), von Hippel explique que selon lui, au paradigme "*manufacturer-active*" qui prescrit que le fournisseur du produit a le rôle d'évaluer les besoins et de développer une idée de produit en réponse, il convient de préférer un nouveau paradigme «*customer-active*» dans lequel le client agit pour développer l'idée de nouveau produit, et prend l'initiative de la transférer à un producteur intéressé.

## **2.2 La question de la source de l'innovation : comment les entreprises captent-elles les rentes liées à l'innovation ?**

Dans l'ouvrage *The Sources of Innovation*, von Hippel (1988) traite, entre autres, de la question de la source de l'innovation, en faisant écho à ces premiers articles relatifs au «*customer active paradigm*». L'auteur cherche à déterminer qui développe réellement les instruments scientifiques qui rencontrent le succès commercial. Dans l'instrumentation scientifique, von Hippel étudie pour différents produits, le premier modèle commercialisé et ses améliorations sur vingt ans. Le résultat de ces observations est que les utilisateurs ont développé 76% des innovations étudiées.

Généralement, dans les cas étudiés, le rôle de l'innovateur dans la diffusion de l'innovation est le suivant :

1. Il perçoit qu'une avancée est nécessaire,
2. Il invente l'instrument,
3. Il construit un prototype,
4. Il démontre la valeur du prototype en l'utilisant,
5. Il diffuse l'information détaillée relative à la fois à la valeur de l'invention et à la façon avec laquelle l'appareil pouvait être dupliqué.

Ensuite, le rôle du fournisseur est généralement le suivant :

1. Il réalise le travail d'ingénierie relatif au produit, sur la base de l'appareil de l'utilisateur pour améliorer sa fiabilité et sa facilité d'exploitation.
2. Il fabrique, réalise le marketing et vend le produit innovant.

Von Hippel (1988) analyse ensuite soixante innovations de procédé de deux types (procédés de fabrication de semi-conducteurs et procédés d'assemblage de circuits imprimés). Il constate que les utilisateurs ont développé plus de 60% des innovations dans ces secteurs. Cette étude montre, par ailleurs, un nombre limité d'activités d'innovation conjointes entre utilisateur et fournisseur. On note que les utilisateurs sont les seuls développeurs des innovations de procédés, relatives à la technique d'utilisation (qui ne demandent pas de nouvel équipement, mais une modification de la façon dont il est exploité). Ainsi l'utilisateur joue souvent le rôle de l'acteur qui reconnaît le besoin, résout le problème au travers d'une invention, construit un prototype et démontre la valeur du produit en utilisation. Le lieu (locus) de presque tout le processus d'innovation est donc centré sur l'utilisateur. Seule la diffusion commerciale est réalisée par le fournisseur. Cela contredit l'approche « manufacturer active » relative au développement de produit nouveau, qui affirme que le fournisseur doit satisfaire un besoin en suivant les différentes étapes du *new product development*.

D'autres études existent attestant d'innovations non-fournisseur, comme celles de Corey (1956). De même, Shaw (1985) étudie le rôle des utilisateurs dans les innovations du secteur de l'instrumentation scientifique en Grande-Bretagne, et en conclut que la majorité des innovations ont été développées, et testées par des utilisateurs. Il observe aussi que des utilisateurs aident souvent les fournisseurs à commercialiser leur innovation, en les assistant pour concevoir, tester et même « marketer » le produit.

### **Echange informel de savoir-faire technique et rente liée à l'innovation :**

Schumpeter (1947) avançait l'idée selon laquelle ceux qui innovaient avec succès sont récompensés par un monopole temporaire sur ce qu'ils ont créés. Ainsi l'innovateur se trouvait récompensé par l'amélioration de sa position sur le marché, et la rente fournie par son innovation. Von Hippel (1988) souligne le fait que les données collectées montrent que les innovations d'un type donné sont développées par des acteurs situés à un niveau particulier de la

filière industrielle étudiée, qui est celui où l'espoir de rente est maximum. Von Hippel (1988) cherche ensuite à explorer de plus près les rentes liées à l'innovation et la façon avec laquelle les innovateurs peuvent les maximiser. Il constate qu'il existe énormément d'échanges informels de *know-how*. Un développement interne requiert de mobiliser un savoir-faire largement informel, que les ingénieurs doivent aller chercher en interne, mais aussi auprès d'experts à l'extérieur de l'entreprise. Il se développe ainsi un réseau informel d'échanges, entre ingénieurs ayant les mêmes centres d'intérêt. En dehors du cas où l'information est vitale pour la firme de l'ingénieur interrogé, celui-ci donne l'information, espérant un retour ultérieur de la part de ses collègues. L'échange informel de know-how peut même survenir entre firmes directement ou non directement concurrentes. Von Hippel focalise ensuite son analyse sur cette coopération entre rivaux. Il cite deux cas célèbres :

- *US steel mini-mill producer* (procédé de production d'acier) : des échanges de savoir-faire ont eu lieu entre rivaux, mais toujours dans l'espoir d'obtenir une contrepartie.
- *Collective invention* décrite par T. Allen(1983), à propos des fonderies anglaises au 19<sup>ème</sup> siècle. Dans ce cas, tous les concurrents ont donné accès à leur savoir-faire propriétaire en même temps.

A l'inverse, von Hippel (1988) note que dans le cas des *PC boards*, les utilisateurs n'ont pas d'incitation à révéler leur innovation, et peuvent même avoir intérêt à la maintenir secrète pour obtenir un avantage concurrentiel. Toutefois, l'auteur note que des transferts ont eu lieu dans un nombre significatifs de cas. Ces transferts (ou non transferts) vers le fournisseur se sont opérés comme suit (par ordre d'importance décroissant) :

1. Grâce à de nombreuses interactions entre les staffs du client et du fournisseur ;
2. Il n'y a pas eu de transfert ;
3. Un utilisateur (pas forcément l'innovateur) a transféré le concept de l'innovation avec la commande d'équipement nouveau requis pour ce concept ;
4. Un utilisateur (pas forcément l'innovateur) a adopté le rôle de fournisseur d'équipement.

### **2.3 Le concept de *lead user*.**

L'article « *Lead user: an Important Source of Novel Product Concept* » (von Hippel, 1986) est le premier dans lequel l'auteur présente le concept de *lead user*, qu'il a lui-même développé. L'auteur

y explique que la précision des études marketing « traditionnelles » dépend de la précision des jugements que portent les utilisateurs sur leurs besoins. Cependant, pour des produits très nouveaux, ou dans des catégories de produits caractérisées par des changements rapides, comme les produits « *high tech* », la plupart des utilisateurs potentiels n'auront pas l'expérience concrète requise pour résoudre le problème et fournir des données précises aux professionnels du marketing. Dans cet article, von Hippel explore le problème, et propose une solution : réaliser des études marketing qui se focalisent sur les *lead users* d'un produit ou d'un procédé. Les *lead users* sont fondamentalement des utilisateurs dont les besoins présents, qu'ils ressentent fortement, se généraliseront sur le marché des mois ou des années plus tard. Puisque les *lead users* sont familiers des conditions qui s'appliqueront dans le futur à la plupart des utilisateurs, ils peuvent servir de laboratoire de prévision des besoins pour les études marketing. De plus, puisque les *lead users* essaient souvent de satisfaire leur besoin, ils peuvent aussi fournir de nouveaux concepts de produit, et des données relatives à leur conception. Dans cet article, von Hippel (1986) explore comment les *lead users* peuvent être identifiés, et comment des perceptions et des préférences de *lead users* peuvent être incorporées dans les études marketing relatives aux produits industriels ou de grande consommation, concernant des besoins émergents de produits, procédés ou services.

Von Hippel (1986) explique que la recherche en *problem solving* prend en compte l'idée que la familiarité avec les utilisations et les attributs de produits existants interfèrent avec la capacité individuelle à concevoir de nouveaux attributs et usages. Les sujets qui utilisent un objet d'une manière familière ont des difficultés à utiliser cet objet de manière innovante. Il en va de même pour les groupes de Recherche dans les entreprises. Ensuite, von Hippel évoque les différentes étapes que doit franchir un utilisateur potentiel quand on lui demande d'évaluer son besoin relatif à un nouveau produit qui lui est proposé. Comme les produits sont des composants de *patterns* (modèles) d'utilisation plus vastes qui impliquent de nombreux produits, et comme un changement concernant un produit peut changer les perceptions de certains produits de ce modèle, les utilisateurs doivent d'abord identifier leur modèle d'utilisation multi-produit, dans lesquels le nouveau produit pourrait jouer un rôle. Ensuite ils doivent évaluer la contribution potentielle du nouveau produit à ces modèles. Ainsi, des changements dans les caractéristiques d'un ordinateur peuvent permettre à un utilisateur de résoudre de nouveaux types de problèmes, s'il fait les changements nécessaires dans certains logiciels. De même, un consommateur adoptant un four à micro-ondes, peut très bien induire des changements dans ses recettes de cuisine, ses pratiques et ses ustensiles de cuisine. Ensuite, l'utilisateur doit inventer ou sélectionner les nouveaux modèles d'utilisation que le nouveau produit proposé rend possible, et évaluer l'utilité

de ce produit dans ceux-ci. Finalement, comme des substituts existent pour de nombreux modèles d'utilisation multi-produits, l'utilisateur doit estimer comment les possibilités offertes par le nouveau produit concurrenceront les options existantes. Cette tâche est très difficile, particulièrement pour les utilisateurs dont la familiarité avec les produits existants interfèrent avec leur capacité à concevoir de nouveaux produits et utilisations.

Von Hippel (1986) explique que ce problème lié à la « familiarité » persiste même si on utilise des techniques marketing sophistiquées comme, par exemple, le *mapping* multi-attributs (le nombre d'attributs étudiés étant forcément limité). Autre exemple, l'intérêt de méthodes recourant à des *focus-groups* (observation des discussions d'un groupe de consommateurs) dépend de la capacité de l'analyste à extraire des entretiens les attributs majeurs du produit. De même, l'analyste doit pouvoir franchir une étape « créative ». En résumé, les professionnels du marketing rencontrent de grandes difficultés à déterminer des besoins de nouveaux produits qui soient éloignés de l'expérience concrète des utilisateurs qu'ils étudient.

### **Prédire la source de l'innovation :**

Comment peut-on prévoir précisément les besoins des utilisateurs pour de nouveaux produits (ou procédés ou services) dans des secteurs qui évoluent rapidement tels que ceux concernés par la *high tech* ? Dans le monde relativement stable de l'acier ou de l'automobile, de nouveaux modèles, ne diffèrent souvent pas radicalement de leurs prédécesseurs. A l'opposé, dans les industries *high tech*, le monde bouge si vite que l'expérience concrète d'utilisateurs ordinaires est souvent déjà obsolète au moment où un produit est développé. Pour de telles industries, von Hippel affirme que le recours des utilisateurs d'un type particulier, les *lead users*, sont essentiels à la compréhension des marchés futurs. Ainsi, von Hippel (1986) définit les *lead users* de produit, procédé ou service nouveau comme des utilisateurs possédant les **deux caractéristiques suivantes** :

1. Ils **rencontrent des besoins qui se généraliseront** sur un marché donné, mais ils les rencontrent des mois ou des années avant la masse du marché. A ce titre, ils sont familiers des conditions qui prévaudront dans le futur pour la plupart des utilisateurs.
2. Ils sont en position de **tirer un bénéfice significatif** de la satisfaction de ces besoins.

La première caractéristique indique qu'un *lead user* possède une expérience particulière que les fournisseurs doivent analyser s'ils veulent comprendre les besoins futurs du cœur du marché ; les utilisateurs « à la pointe de la tendance » (*leading edge*), existent parce que les technologies, les produits, ou les goûts se diffusent à travers la société, souvent pendant des années, plutôt que d'en impacter chaque membres simultanément (Rogers, 1995). La deuxième caractéristique des *lead users*, est une application directe de l'hypothèse avancée par von Hippel (1986) selon laquelle des utilisateurs qui espèrent des rentes élevées de la satisfaction d'un besoin, vont essayer de trouver une solution. Les *lead users* seraient donc en mesure de fournir une description perspicace de besoins, et peut-être même de solutions.

Pour justifier de l'existence et de l'utilité quant au processus d'innovation, des *lead users*, von Hippel (1994) explique qu'elles résultent principalement de deux éléments : tout d'abord la « *stickiness* » (ou « adhérence ») de l'information nécessaire au développement de l'innovation, notion que nous explicitons ci-après, et d'autre part, l'hétérogénéité des besoins des utilisateurs sur laquelle l'auteur revient à propos des systèmes d'information des bibliothèques australiennes et du logiciel Apache.

## 2.4 L'adhérence (« *stickiness* ») de l'information :

Von Hippel (1994) décrit en détail la notion d' « adhérence » de l'information, qu'il a développée dans les années 1980, et qui selon lui explique en grande partie pourquoi les utilisateurs sont dans une position plus favorable que les fournisseurs pour innover. Von Hippel (2005) s'inscrivant dans le paradigme du *problem solving*, part d'un constat simple : « *The goal of Product Development is to create a solution that will satisfy needs of real users within real context of use* ». Pour résoudre un problème, l'information concernant ce problème (les besoins et le contexte de l'utilisation), et les capacités de résolutions (solutions possibles) qui s'y rapportent doivent être réunies en un seul et même lieu. Cela ne pose pas de difficulté si le transfert de ces deux formes d'information peut s'effectuer à moindre coût ; mais lorsque ce coût de transfert devient important, la connaissance est alors « adhérente » (*sticky*) au sens de von Hippel (1994). L'« adhérence » d'une unité d'information est définie comme la dépense (« *expenditure* ») requise pour transférer cette unité d'information dans un lieu donné sous une forme utilisable par celui qui a besoin de cette information pour résoudre le problème (*information seeker*). Von Hippel (1994) définit l'information « adhérente » comme l'information qui est coûteuse à acquérir, transférer, et utiliser dans un autre lieu.

Pour le développeur d'un produit innovant, l'identification des besoins des utilisateurs est une étape essentielle, qui est contrainte par les coûts d'acquisition de l'information pertinente. Les développeurs de produits rencontrent en général rapidement des problèmes dûs au fait que les utilisateurs détiennent une fraction essentielle, mais souvent *adhérente* de l'information nécessaire à l'accomplissement de leur tâche. L'information requise par ceux qui sont chargés de résoudre des problèmes techniques, est souvent fortement *adhérente* pour des raisons qui peuvent être liées à ses attributs, telles que la façon dont elle est encodée (Nelson 1982, Pavitt 1987, Rosenberg, 1982) ; l'« adhérence » peut aussi provenir des limites des capacités d'absorption (*absorptive capacities* ; Cohen and Levinthal, 1990) des *information seekers*. De même, des personnels spécialisés tels que les « *technological gatekeepers* » (R. Allen, 1984) et les structures organisationnelles spécialisées telles que les groupes de transfert d'information (Katz et R. Allen, 1988) peuvent affecter les coûts de transfert d'information entre et à l'intérieur des organisations.

Pour ces raisons, les informations relatives aux besoins des utilisateurs sont fréquemment très *adhérentes*, ce qui signifie que l'information requise par les développeurs de produits nouveaux ne peut être échangée qu'à des coûts élevés. Pour von Hippel, le caractère « *sticky* » de certaines informations relatives au développement de l'innovation, explique que certaines innovations ne peuvent être conçues que par des utilisateurs, et plus particulièrement des *lead users*.

## 2.5 Lien entre la théorie de la diffusion et le concept de *lead user*.

Pour von Hippel (1986, 2005), il existe des utilisateurs qui présentent des besoins en avance sur une demande générale, parce que les nouvelles technologies, les nouveaux produits, ou les nouveaux procédés qui se diffusent à travers la société, n'impactent pas tous ses membres simultanément (Rogers, 1995). Mansfield (1972) étudiant la diffusion d'innovations dans l'industrie, trouva que 75% de celles-ci mirent plus de 25 ans à se diffuser dans les grandes entreprises. Il serait donc possible de trouver systématiquement des utilisateurs situés « à la pointe » ou « en avance de phase » par rapport au marché. Par ailleurs, les utilisateurs se distinguent les uns des autres par l'ampleur des bénéfices qu'ils peuvent espérer retirer de l'utilisation d'un nouveau produit. Plus il est élevé, plus ils auront tendance à faire des efforts pour l'obtenir (Schmookler, 1966), et à consacrer des ressources pour le comprendre. Ainsi, on peut estimer logiquement que parmi les utilisateurs situés « à la pointe du marché », il existe un sous-ensemble d'utilisateurs qui devrait avoir la plus riche expérience concrète relative au nouveau besoin étudié.

## 2.6 Le concept de *lead user* comparé à d'autres concepts de la littérature :

Les principaux travaux dans le champ de l'innovation traitent généralement de deux problématiques : l'analyse des processus d'innovation ou l'analyse du processus de diffusion de l'innovation. Dès ses premières recherches, Von Hippel (1977) étudiant les utilisateurs à l'origine d'idées nouvelles, s'intéresse à la phase amont du processus d'innovation. Il cherche à démontrer que les consommateurs peuvent être sources d'idées nouvelles au même titre que les inventeurs tels qu'on peut les rencontrer dans les entreprises. Pour sa part, Rogers (1983), dans ses travaux relatifs à la diffusion de l'innovation, parle des *innovateurs* qui sont les premiers à adopter (« *early adopters* ») le nouveau produit. Il ne faut donc pas confondre les *early adopters*, avec les *lead users* qui, eux, agissent quand il n'y a rien à adopter, parce que l'un de leurs besoins est mal comblé. Pour résumer, les *lead users* ne sont pas des *innovateurs* au sens de Rogers, mais des *inventeurs* (au sens sociologique du terme, Alter 2000), qui, de plus, anticipent les attentes de la majorité des autres utilisateurs.

Par ailleurs, il ne faut pas confondre le concepts de *lead user* et celui de leader d'opinion. Le leader d'opinion est caractérisé par sa capacité à communiquer des informations à son groupe d'appartenance et à en influencer les membres (Ben Miled et Le Louarn, 1994). Dans le cadre de la diffusion de l'innovation (Rogers, 1983), le leader d'opinion est un relais d'information et entraîne la majorité des consommateurs à adopter le nouveau produit. Le *lead user*, par l'expression d'attentes en avance sur son temps, présente quant à lui une autre forme de leadership. On peut le qualifier de leader seulement en termes d'idées nouvelles. Toutefois, Morrison (1994) montre que les *lead users* ont tendance à communiquer leur expérience et à influencer les autres, c'est-à-dire à avoir certains comportements similaires à ceux des leaders d'opinion. Ce résultat permet de supposer que les *lead users* jouent un rôle intéressant dans la diffusion de l'innovation.

## 2.7 Le concept de *lead user* dans la littérature :

Nous observons tout d'abord que différents résultats issus de la recherche sur l'innovation, préalables à la définition de la notion de *lead user*, ont mis en évidence des éléments relatifs à le rôle des utilisateurs, comme source d'un grand nombre d'innovations. Ainsi, Enos (1962), Freeman (1968), et Shaw (1985) montrent que les utilisateurs sont, plus souvent que les fournisseurs, à l'origine des succès commerciaux de nouveaux produits.

Par la suite, on constate que peu de travaux ont porté sur le concept de *lead user* (Becheur & Gollety, 2006). Il est intéressant de constater que ce concept n'a jusqu'alors jamais été réellement contesté : l'idée qu'une partie significative des innovations ait pour origine certains utilisateurs, et que ceux-ci présentent des caractéristiques qui en font des *lead users* au sens de von Hippel (1986), semble admise, même si elle ne donne pas lieu à une exploration académique particulière (Le Masson, Weil & Hatchuel, 2006). C'est d'autant plus remarquable que von Hippel publie très régulièrement, seul ou avec des co-auteurs, des articles dans des revues de référence en science de gestion, depuis une vingtaine d'années. Ses publications s'enchainent affinant et infléchissant ses concepts de départ. Si le concept de *lead user*, en particulier, semble avoir trouvé un écho favorable dans les pratiques marketing de certaines entreprises, notamment aux Etats-Unis, peu de recherches académiques s'y sont intéressées. Et pourtant l'intérêt du concept de *lead user* réside dans le fait de pouvoir identifier ces utilisateurs par des caractéristiques précises. Les études disponibles à ce jour (Voss, 1985, Morrison, 1994, Béji-Bécheur, 1998) sur ce sujet se fondent toutes sur la définition de von Hippel.

Voss (1985) et Béji-Bécheur (1998) montrent que les utilisateurs leaders (« *lead user* » pouvant aussi être traduit par « utilisateur pionnier ») ont tendance à avoir un niveau d'expertise, dans leur domaine, supérieur à la moyenne. Ce sont des « experts généralistes » dans le domaine sans être pour autant des spécialistes pointus. Ils possèdent une culture large du domaine, et sont capables de faire référence à l'histoire, ainsi qu'à l'environnement socioculturel et économique pour enrichir et développer leur réflexion. A l'inverse, l'« expert spécialiste » a un cadre de réflexion trop restreint à son domaine de spécialité pour avoir une vision globale du produit dans son environnement. Pour Béji-Bécheur (1998), les *lead users* sont familiers de la nouveauté dans leur domaine. Ils suivent les évolutions du ou des domaines qui les intéressent. En outre, ils ont tendance à développer des idées nouvelles. Ils imaginent des solutions pour améliorer le produit ou réalisent eux-mêmes des modifications sur leur produit. Par ailleurs, Morrison (1994), observe que l'innovation provenant des utilisateurs tend à se concentrer parmi les *utilisateurs pionniers* de ces produits.

En revanche, ces études font apparaître que les *lead users* ne sont pas forcément les premiers acheteurs de tous les produits nouveaux, autrement dit des adopteurs précoces au sens de Rogers (1983). La nouveauté en soi ne les intéresse pas. Ils n'achètent pas un nouveau produit parce qu'il est nouveau mais parce qu'il correspond à un besoin précis. Morrison (1994) montre que les *lead users* ont tendance à communiquer leur expérience et à influencer les autres, c'est-à-dire à avoir

certaines comportements similaires à ceux des leaders d'opinion. Ce résultat permet de supposer que les *lead users* jouent un rôle intéressant dans la diffusion de l'innovation. De l'étude comparée des *lead users* et des leaders d'opinion, ce même auteur fait cependant ressortir que les *lead users* ont plus d'expérience liée aux innovations en général que les leaders d'opinion et, que leur niveau de connaissance de l'innovation est plus élevé que celui des leaders d'opinion. Ce résultat indique que si les utilisateurs leaders ont tendance à être leaders d'opinion, l'inverse n'est pas vrai. Ainsi, un leader d'opinion peut même représenter un groupe réfractaire à l'innovation.

Si les travaux que nous venons de présenter permettent de mieux cerner les caractéristiques des *lead users*, ils reposent tous sur la définition du concept proposée par von Hippel (1986). Cette définition et ce concept n'ont pas été remis en cause, ni réellement critiqués par la littérature depuis. Toutefois, pour certains auteurs, le manque de précision relatif aux fondements de la notion de *lead user*, et la difficulté à développer un outil de sélection de ces utilisateurs-là, invitent à affiner les déterminants de ce concept (Becheur & Gollety, 2006). En définitive, la littérature relative à l'innovation par les utilisateurs, depuis vingt ans (von Hippel, 1986, 1988, 2005), a eu pour principal objet de démontrer l'importance du rôle des *lead users* dans la conception d'innovation, et les différentes formes que pouvait prendre leur participation aux processus d'innovation.

## 2.8 Une approche « universelle » ?

L'idée de von Hippel était, dans ses premiers travaux, que son approche concernait surtout les univers high-tech où les produits évoluent très rapidement. Herstatt et von Hippel (1992) essayent de montrer que cette approche reste valable dans les univers *low tech*, au travers du cas de l'entreprise Hilti, entreprise de fabrication de produits et de matériels utilisés en construction. Des concepts de produits nouveaux (destinés à la fixation de tuyaux) ont été développés lors d'un atelier de trois jours organisé par Hilti, auxquels ont assisté les *lead users* et le personnel de développement et de production de Hilti. Ces ateliers ont été l'occasion de mettre en œuvre ce que von Hippel (1986) appelle la méthode « *lead user* ». En effet, von Hippel (1986), à partir de son approche *lead user*, a défini une « méthode *lead user* » que nous décrivons ici de manière abrégée, et qui se décline en une série d'étapes successives :

1. Identifier une tendance technique ou de marché.

2. Identifier des *lead users* qui sont à la pointe de cette tendance, et qui espèrent retirer un profit élevé de la satisfaction de leur besoin,
3. Analyser les données relatives au besoin utilisateur : des *lead users* peuvent avoir investi sur une solution à leurs besoins ; parfois ils vont mettre en œuvre des produits existants d'une manière inédite, parfois ils vont développer des produits complètement nouveaux pour répondre à leur besoin.
4. Projeter les données *lead user* sur l'ensemble du marché.

Dans le cas de l'entreprise Hilti, la phase 3 a consisté à monter des ateliers permettant de concevoir et réaliser en temps réel une innovation. Hilti a donc mis en œuvre cette méthode avec succès, puisqu'il a en a résulté une conception deux fois plus rapide, mais aussi moins coûteuse, qu'avec les méthodes habituellement employées par cette entreprise (Herstatt et von Hippel, 1992).

Ainsi, en avançant que l'approche et la méthode *lead user* s'applique tout aussi bien à des produits *low tech*, voire mieux puisque des ateliers de conception permettent de concevoir et réaliser en temps réel une innovation grâce à des *lead users*, Herstatt et von Hippel (1992) semblent vouloir montrer que l'approche *lead user* présente un caractère universel.

## 2.9 L'approche *lead user* pour générer des innovations radicales :

L'article « Creating Breakthroughs at 3M » de von Hippel, Thomke, et Sonnack (1999), adresse une problématique qui est devenue, à partir des années 1990, cruciale pour toute entreprise technologique, la génération régulière d'innovations radicales (« *breakthrough* »). La crise du début des années 90 a eu comme conséquence une nécessaire re-focalisation des entreprises sur des enjeux à court terme visant tout simplement la survie de ces dernières, limitant dès lors les investissements en recherche et développement des innovations. Or, depuis la fin des années 90, la problématique est de nouveau lancée sur la nécessité d'acquérir les mécanismes d'une croissance solide et durable ; la question du développement de l'innovation, et notamment d'innovation de rupture (*breakthroughs* et non plus des innovations incrémentales) reprend dès lors tout son sens, comme condition sine qua non pour maintenir la position de l'entreprise sur le marché sur le long terme. C'est dans ce contexte que les auteurs, von Hippel, Thomke et Sonnack (1999) se pose la question fondamentale : Comment rationaliser le processus de conception des

innovations de rupture ? Ils cherchent alors à savoir dans quelle mesure les *lead users* constituent une source de conception des innovations de rupture, et à déterminer les modalités d'application de la méthode *lead user*.

Pour ces auteurs, les entreprises ont des difficultés à développer des innovations radicales, car elles ne disposent d'aucun système efficace pour guider leurs efforts. Ils poursuivent en expliquant que l'on apprend en général aux équipes de développement à rassembler des informations relatives aux utilisateurs qui sont au centre de leur marché cible. Ensuite, elles comptent sur leur propre puissance créatrice pour imaginer de nouvelles idées. Le processus *lead user* adopte une approche différente, car il est conçu pour rassembler des informations à la fois sur les besoins et sur les solutions des principaux « bords de la cible », et des marchés qui font face à des problèmes semblables sous une forme extrême. Les auteurs décrivent les processus d'innovation intégrant l'approche *lead user*, chez 3M de la manière suivante :

« Les *lead users* aident souvent les équipes-projet à améliorer leur compréhension de la nature de l'innovation radicale qu'ils recherchent. [...] Il est rare qu'un fournisseur adopte une innovation *lead user* telle quelle, mais les nouveaux concepts de produit sont souvent basés sur des informations recueillies auprès de différents *lead users* et de développeurs internes. [...] 8 des 55 divisions de 3M utilisent maintenant la méthode *lead user*. [...] Cette méthode incite des équipes pluridisciplinaires à travailler avec des clients à la pointe d'une tendance. » (von Hippel, Thomke et Sonnack (1999).

La conclusion de cet article, est que la mise en œuvre de la méthode *lead user* chez 3M, a permis de générer des innovations radicales plus rapidement, mais cependant d'une manière plus onéreuse, que les méthodes habituelles. Ce dernier point vient en contradiction avec les observations faites chez Hilti, qui ne concernaient pas des innovations radicales. En conclusion, von Hippel, Thomke et Sonnack (1999) affirment que la méthode *lead user* est susceptible de produire de manière régulière des innovations radicales.

## 2.10 La notion de *toolkit* : définition et lien avec l'innovation utilisateur

### 2.10.1 Introduction :

Dans le domaine du logiciel, les *toolkits* sont des environnements de développement fournissant des outils de développement de logiciels ou de modules complémentaires. Ces *toolkits* peuvent être utilisés en interne par des éditeurs de logiciel, par des sociétés de service ou encore par les utilisateurs finaux eux-mêmes. La notion de *toolkit* est utilisée par von Hippel (2001, 2002), mais il n'est pas le seul auteur à s'y être intéressé. Mikkola (2003, 2006), entre autres, montre que l'approche *toolkit* permet la personnalisation des solutions et la *mass customization*. Mais von Hippel l'a adoptée comme un des éléments clefs de son approche *lead user*, ainsi que nous allons le voir.

Les travaux de von Hippel (1994,1998) sur l'innovation utilisateur, l'ont amené à s'intéresser à une nouvelle approche, fondée sur l'utilisation de *user toolkit* (« *toolkit for user innovation* » qui peut être traduit par *boîtes à outils* ou *trousses d'appui* à l'innovation utilisateur), pas uniquement dans le domaine du logiciel, même si c'est dans ce domaine que son application est la plus répandue. Cet intérêt porté par von Hippel aux *toolkits* est logique parce que ceux-ci renforcent la capacité des utilisateurs à concevoir leurs propres solutions, et donc à innover. Ainsi de l'idée qui consiste à observer comment les *utilisateurs* innovent et à intégrer leurs meilleures idées dans les innovations, von Hippel (1994) est passé à celle qui consiste à les équiper de manière à ce qu'il soit plus facile pour eux d'innover à la place du fournisseur. Cette idée a émergé dans les années 1980, dans le domaine des logiciels de conception assistée par ordinateur (CAO). Par exemple, les *trousses d'appui* à l'innovation client sont apparues dans l'univers de la conception et de la fabrication sur mesure de circuits intégrés. Dans ce secteur, les produits étaient hautement complexes. Par conséquent, il en coûtait très cher aux fournisseurs de ne pas comprendre rapidement et précisément ce qu'un acheteur donné avait en tête. Pour contourner le fait qu'ils avaient de la difficulté à définir correctement les besoins de leurs clients, les fournisseurs ont commencé à les équiper de *trousses d'appui*. Ainsi, les acheteurs pouvaient faire eux-mêmes la conception des produits qu'ils souhaitaient acquérir du fournisseur. L'introduction des *trousses* a d'ailleurs permis de réduire des deux tiers les délais de développement dans le secteur des semi-conducteurs.

Ainsi, l'approche *toolkit* pour l'innovation est complémentaire de l'approche *lead user* : certains des *utilisateurs* utilisant un *toolkit* seront des *lead users*. Conscient de l'intérêt d'une approche *toolkit*, l'éditeur de logiciel Stata Corporation vend un package logiciel pour réaliser des analyses

statistiques complexes qui la prend en compte. Ce package offre les fonctions d'un *toolkit* et Stata encourage les clients à créer et partager de nouveaux codes du logiciel pour exécuter de nouvelles techniques statistiques. L'entreprise sélectionne ensuite les développements d'utilisateur dignes d'intérêt pour de nombreux utilisateurs, les adapte et les incorpore dans les versions ultérieures de ses produits. Enfin, on peut observer que le concept de *boîte à outils* pour l'utilisateur-innovateur se présente comme une stratégie subtile permettant un bon équilibre entre diversité et coût, tout en favorisant l'apprentissage des utilisateurs.

Selon von Hippel (interview Cefrio, 2002), trois signes indiquent qu'une entreprise pourrait et devrait sans doute se doter de *trousses d'appui* à l'innovation-client :

1. Ses segments de marché diminuent et les clients exigent de plus en plus de produits faits sur mesure. L'entreprise cherche à s'adapter à ces changements, mais ses coûts augmentent sans pouvoir être transférés à ses clients.
2. Un grand nombre d'itérations sont requises avant que le client puisse obtenir un produit satisfaisant.
3. L'entreprise ou ses concurrents utilisent des logiciels de simulation ou de prototypage en interne.
4. Les procédés de fabrication de l'entreprise sont adaptables et permettent de produire des articles sur mesure.

### 2.10.2 Le transfert de certaines activités de conception chez l'utilisateur :

L'approche *toolkit* peut entraîner un transfert des capacités d'innovation, le fournisseur procurant une *boîte à outils* à son client de manière à ce que ce dernier puisse concevoir et développer une version du produit qui soit parfaitement adaptée à ses besoins. Cela transfère le lieu de l'interface fournisseur-client, et le processus itératif d'essai-erreur est alors supporté par le client. Il en résulte un gain de temps et d'efficacité car le fournisseur n'a, idéalement, plus à comprendre en détail les besoins de clients. Ainsi, l'approche *toolkit*, entraînant un transfert de pouvoir à l'utilisateur comporte certains risques : « Il faut donc en tenir compte au moment de prendre la décision d'aller de l'avant ou non. Certains clients pourraient utiliser votre *toolkit* contre vous et se transformer en rivaux. Les avantages et les désavantages de cette approche doivent être soupesés » (von Hippel, 2002).

Prenons l'exemple des semi-conducteurs : traditionnellement les fournisseurs de semi-conducteurs personnalisés réalisaient toutes les tâches de conception de la puce, guidés seulement par les spécifications fonctionnelles (« *need specifications* ») des utilisateurs. La partition de la tâche de conception de ces semi-conducteurs en sous-tâches d'une part liées à la solution et d'autre part liées au besoin, devint possible parce que la conception d'éléments fondamentaux des puces, tels que ses transistors, devint standard pour tous les concepts de circuits numériques personnalisés. Cette sous-tâche requerrait un accès poussé à l'information *adhérente* liée à la solution du fournisseur, mais ne requerrait pas d'information détaillée relative aux besoins des utilisateurs. On a ainsi pu assigner à la puce, chez le fournisseur, des ingénieurs de fabrication et de conception. Parallèlement, on a observé que la sous-tâche d'interconnexion des éléments de circuits numériques standards dans un circuit intégré en fonctionnement, demandait seulement des informations *adhérentes* liées au besoin concernant la fonction de la puce (par exemple, si elle devait constituer le microprocesseur d'un ordinateur ou la puce vocale d'un chien robot). Ces sous-tâches de câblage du circuit furent ainsi assignées aux utilisateurs. En résumé, ce nouveau type de puce nommé « *gate array* », disposait d'une nouvelle architecture créée spécifiquement pour séparer les tâches de *problem solving* (requérant l'accès à des informations *adhérentes* détenues par le fournisseur), de celles réclamant l'accès à l'information *adhérente* des utilisateurs.

Quand un circuit intégré est entièrement conçu par des fournisseurs, les utilisateurs ne peuvent s'engager dans le processus itératif décrit plus haut, tant qu'une puce n'a pas été complètement conçue et livrée par le fournisseur. Dans une telle situation, la prise en compte des modifications réclamées par les utilisateurs peut s'avérer très coûteuse. A l'inverse, des utilisateurs équipés d'un *toolkit*, peuvent apprendre à identifier et à corriger très tôt des erreurs de conception, ce qui peut contribuer à une réduction spectaculaire des coûts de conception.

### 2.10.3 Le développement des *toolkits* par le fournisseur :

Les *toolkits* constituent donc une façon de transférer les capacités de conception aux utilisateurs, mais ceux-ci n'étant pas des spécialistes de la conception, comment faire en sorte qu'ils créent des concepts complexes et, dans le cas des produits manufacturés, susceptibles d'être produits industriellement ? Pour von Hippel (1994, 1998), les fournisseurs doivent résoudre ce problème en proposant aux utilisateurs des *toolkits* qui leur permettent de se concentrer que sur les tâches de conception qui permettent d'adapter précisément l'innovation à leur besoin. Les fournisseurs proposant des *toolkits* sont alors, dans l'idéal, libérés de la nécessité de connaître le détail des

besoins de leurs clients. Mais, pour cela, le fournisseur doit connaître l'« espace de solution » dont ses clients ont besoin pour être capable de concevoir des innovations. Concrètement, les fournisseurs peuvent souvent créer un premier *toolkit* en modifiant leurs outils de conception internes pour les rendre plus accessibles, et les proposer comme *toolkits* de première génération. L'utilisation de tels *toolkits* conduit les clients à se heurter aux limites de l'espace de solution qu'ils décrivent, et les incitent ensuite à réclamer des améliorations au fournisseur. Les fournisseurs peuvent aussi attendre que les utilisateurs améliorent eux-mêmes leur *toolkit*. Fondamentalement, le concept de *toolkits* est lié à la notion de modularité, puisque la conception d'une architecture modulaire conduit au « découpage » d'un produit en modules indépendants. Une telle architecture peut alors permettre de distinguer entre modules génériques et modules personnalisables (von Hippel, 1990). Ces derniers peuvent alors, devenir les *toolkits* utilisables par les utilisateurs. La tâche de conception du fournisseur se prolonge alors, qui l'amène à concevoir des modules de personnalisation exploitables par les utilisateurs, les *user toolkits*. Ainsi, les *toolkits* peuvent concourir à résoudre le problème de la différenciation et de la personnalisation des produits, à bas coût.

Thomke et von Hippel (2002) affirment que les *user toolkits* doivent posséder les caractéristiques suivantes :

- Ils doivent permettre un **apprentissage de type « learning by doing »** : simulation par ordinateur ou, s'il faut «toucher» l'objet, procédés de prototypage rapide ;
- Ils doivent posséder une interface conviviale et **ne pas requérir l'apprentissage d'un nouveau langage** ;
- Dans le cas des produits manufacturés, les outils doivent intégrer des **informations relatives à l'industrialisation** de leur production.

Ces caractéristiques éclairent les raisons pour lesquelles de tels outils sont largement répandus dans le secteur du logiciel, où le problème de l'industrialisation de la production ne se pose pas et, où les développeurs peuvent utiliser des langages informatiques standards.

## 2.11 La contribution des communautés d'utilisateurs :

Quand les besoins d'utilisateurs sont hétérogènes et quand l'information qu'ils possèdent est particulièrement *adhérente*, von Hippel (2005) affirme que les activités de développement de

produit sont susceptibles d'être largement distribuées parmi les utilisateurs. Le principe de l'innovation utilisateur distribuée repose sur l'idée selon laquelle que chaque utilisateur présente une double spécificité : tout d'abord, il ressent des besoins spécifiques, et ensuite, il est doté d'une capacité (de par ses compétences, ses moyens techniques ou financiers, ses centres d'intérêt, etc.) à innover qui, elle aussi, lui est propre. Cette double spécificité fait qu'il aura tendance à innover d'une manière différente des autres. Dans certains cas, ces apports peuvent alors contribuer à un processus d'innovation qui les englobe. Il peut alors se former des *user communities* ou communauté d'utilisateurs (von Hippel, 2001), qui favorisent l'apprentissage et le partage de savoirs, comme la communauté Open Source dans le logiciel (Raymond, 1999).

Franke et von Hippel (1999) et von Hippel (2001) affirment que certaines conditions doivent être satisfaites pour qu'une communauté d'utilisateurs fonctionne :

1. Il existe **nombre significatif d'utilisateurs motivés** par l'innovation : cette première condition est celle de l'incitation à innover, et donc de la perception par les utilisateurs des bénéfices futurs qui pourront être retirés de l'utilisation de cette innovation. Dans cette première condition, on retrouve également le niveau de compétence minimal que les utilisateurs doivent détenir pour être capable d'innover ;

2. Les **innovateurs acceptent de révéler** le contenu de leurs apports : cette deuxième condition est nécessaire pour que ce type de système fonctionne sur le long terme ;

3. Le **coût de diffusion de l'information est très faible** : l'usage des technologies de l'information est à cet égard crucial. Mais pour certaines communautés (comme celles des sportifs), les rencontres entre utilisateurs à l'occasion de grands événements assurent ce rôle de partage des connaissances à faible coût.

En conclusion sur ce point, nous observons que la littérature sur l'innovation par les utilisateurs souligne l'importance des communautés d'utilisateurs dans la dynamique d'innovation (von Hippel, 2005), notamment dans l'industrie logicielle. Celle-ci fournit des exemples nombreux de communauté d'utilisateurs contribuant à l'innovation ou à l'amélioration des produits, que ce soit les communautés Open Source dans l'univers des logiciels éponymes, ou les *user groups* (ou groupes d'utilisateurs) dans l'univers des logiciels propriétaires (Raymond, 1999 ; Franke et von Hippel, 2003).

## 2.12 Conclusion sur l'approche *user innovation*:

Le logiciel apparaît effectivement comme un champ d'application privilégié de l'innovation par les utilisateurs. Cela résulte d'abord de son immatérialité, qui permet de dupliquer un logiciel à un coût très faible, sans passer par une phase de production industrielle. Cela résulte ensuite de la modularité de son architecture, qui permet d'éclater les tâches de conception et de développement, les deux étant intimement liés dans ce cas, entre différents acteurs. Mais il n'en demeure pas moins que, dans le cas de l'innovation de rupture, l'appropriation de concepts et de technologies radicalement nouveaux par les utilisateurs, concepts et technologies développés par des petits éditeurs de logiciel spécialisés, suscite des interrogations. Dans le cas du logiciel, comment se conjuguent ces deux éléments contradictoires, terrain favorable à l'innovation utilisateur et difficulté d'apprentissage inhérente à toute innovation radicale proposée par un fournisseur ? Dans quelle mesure les *lead users* constituent-ils une source de conception d'innovations de rupture ? Quel rôle peut concrètement jouer un *lead user* dans le processus d'innovation d'un logiciel radicalement innovant ? Comment les éditeurs de logiciel doivent-ils prendre en compte la propension de ce secteur à l'innovation utilisateur, pour gérer le processus d'innovation ?

En ce qui concerne les notions de *toolkits* et de communauté d'utilisateurs, l'étude d'une innovation logicielle radicale nous conduira aussi à déterminer les conditions dans lesquelles des *toolkits* destinés aux utilisateurs et de telles communautés peuvent émerger.

## 3. LA PERSPECTIVE RELATIONNELLE RETENUE PAR LE MARKETING INTERORGANISATIONNEL :

### 3.1 Introduction :

La plupart des travaux de recherche en marketing industriel antérieurs aux années 1980, se sont orientés selon deux voies (Bonoma et Johnston, 1976 ; Ford, 1980) : l'une s'est focalisée sur le fonctionnement du centre d'achat des clients pour comprendre les décisions d'achat et des facteurs qui affectent le choix des fournisseurs ; l'autre s'est intéressée à la décision marketing du point de vue du fournisseur, et a étudié l'effet des différents éléments du marketing-mix sur les marchés industriels. Ces approches étaient des transpositions directes de celles utilisées en marketing des biens de grande consommation. Le champ du marketing industriel, et plus

généralement celui du marketing interorganisationnel, a ensuite évolué à partir des années 1980, en adoptant une perspective relationnelle (Arndt 1979, Dwyer, Schurr et Oh 1987). Fondamentalement, cette perspective met en cause une approche centrée sur le fournisseur, l'approche diffusionniste (Rogers, 1962). Elle est semblable en cela à l'approche *lead user*, qui met, elle, en cause un modèle dans lequel le fournisseur assume l'entière responsabilité du processus de conception.

Selon cette perspective, c'est l'étude des relations entre acteurs organisationnels, plutôt que celle des transactions, qui permet l'étude des échanges interorganisationnels. Lambe et al. (2000) indiquent ainsi que si le marketing reste préoccupé d'échange (Bagozzi ; 1975), il est surtout préoccupé de la forme relationnelle que prend cet échange. Dans cette perspective, c'est la nature de la relation qui permet à un échange de créer de la valeur, et qui gouverne cet échange.

Plus précisément, des chercheurs, au sein du groupe de recherche européen IMP (*International Marketing and Purchasing*) ont mis en évidence des différences fondamentales entre marchés industriels et marchés dit de grande consommation, qui les ont amené à « observer *l'interaction* [entre le fournisseur et le client], à adopter une *perspective* long terme, à prendre en compte la *stabilité* de la relation [entre fournisseur et client], à reconnaître le caractère *actif* des deux acteurs, et enfin à ne pas réduire la stratégie du fournisseur à la simple manipulation des leviers d'action du marketing-mix » (Hakansson, 1982 ; Valla, 1987 ; Salle et Sylvestre, 1992).

## 3.2 Les travaux du groupe IMP :

### 3.2.1 La perspective relationnelle :

Les chercheurs du groupe de recherches européen IMP (*International Marketing and Purchasing*), adoptent donc cette perspective relationnelle ou « approche interaction » (Hakansson, 1982), et proposent l'utilisation d'une nouvelle unité d'analyse pour étudier les activités marketing d'une entreprise (Ford et al. 1998). La « relation » remplace la « transaction » comme objet d'étude dans les échanges. Ils se sont ainsi focalisés sur l'analyse approfondie du déroulement de relations entre des fournisseurs et leurs clients, pour en tirer des conclusions susceptibles d'être mises en œuvre par les entreprises sous forme de méthodes. Le principal résultat théorique de ces travaux est la position centrale accordée à la relation entre le fournisseur et le client, comme lieu où se construit, se déroule et se régule l'échange marchand, et donc comme facteur explicatif essentiel

de la dynamique des marchés industriels (Hakansson, 1982). A cela, ces chercheurs ajoutent une proposition relative à une nouvelle conception du processus de ces échanges. Ceux-ci ne sont pas le fait d'une partie agissant sur l'autre - modèle classique de type stimulus-réponse (Bonoma et Jonhston (1978) - mais de deux parties agissantes, fournisseur et client étant engagés dans une relation interactive.

### 3.2.2 De la dyade au réseau :

Toutefois, le groupe IMP ne s'est pas uniquement focalisé sur la dyade client/fournisseur, car une partie de la compréhension des échanges interorganisationnels passe par la prise en compte d'autres acteurs connectés directement ou indirectement à cette dyade. Pour Hakansson et Snehota (1989), une entreprise peut être vue comme une île et la source de son avantage concurrentiel réside dans sa capacité à relier des entreprises (donc des ressources) à d'autres. Cette perspective a conduit les chercheurs du groupe IMP à une deuxième conceptualisation celle des "marchés comme réseaux" (Johanson et Mattson 1994 ; Hakansson et Snehota 1995), conceptualisation appelée approche « réseau ». Il ne s'agit pas ici de remplacer une conceptualisation (la relation entre deux acteurs en interaction) par une autre (le réseau de relations entre plusieurs acteurs), mais de réenrichir la relation dans un contexte élargi. Le réseau, dans la perspective du groupe IMP, fournit donc un niveau d'analyse complémentaire à celui de la relation (Hakansson et Snehota, 1989). Fondamentalement, l'intérêt de ce type d'approche « élargie » est de resituer l'entreprise non plus dans un « *faceless environment* », mais dans un ensemble de relations avec des acteurs dénommés ; ce sont les relations avec ces acteurs qui vont permettre d'accéder à des ressources détenues par d'autres acteurs. (Hakansson et Snehota, 1989).

### 3.2.3 Le modèle d'interaction :

Retenant à la fois les perspectives relationnelle et « réseau », les chercheurs du groupe IMP ont bâti un « modèle d'interaction » (Hakansson, 1982), modèle qui se concentre sur la relation fournisseur/client appréhendée comme un processus d'interaction constitué par des échanges entre un groupe d'achat chez le client et un groupe de vente chez le vendeur. Le « modèle d'interaction » (Hakansson, 1982 ; Valla 1987) vise à intégrer tous les éléments caractérisant la relation fournisseur/client. Il permet aussi d'analyser la relation entre deux fournisseurs dans le cas du développement conjoint d'une offre (Hakansson, 1987). Le modèle décrit et relie, donc,

les éléments à prendre en compte pour comprendre et expliquer les échanges fournisseur-client (ou entre fournisseurs), comme l'indique la figure 3 (p 92). Ces éléments sont (Hakansson, 1982) :

(a) Le **processus d'interaction** : Le *processus d'interaction* rend compte du contenu, du déroulement et de l'évolution des échanges entre deux organisations. Il est rythmé par une succession d'épisodes, tels que transactions commerciales, crises dues à des litiges techniques ou logistiques, phases de développement de nouveau produit, etc. La perspective retenue ici se distingue donc d'une perspective transactionnelle, i.e. focalisée sur les transactions entre fournisseur et client. A propos du processus d'interaction, quatre points sont à souligner :

- Ce processus est appréhendé selon les perspectives court-terme et long-terme.
- Les épisodes individuels renvoient aux échanges sur le court-terme,
- Dans ces épisodes, quatre types d'échanges sont distingués : échanges de produits ou services, échanges financiers, échanges d'informations, et échanges sociaux.
- La relation à long-terme résulte de cet ensemble d'épisodes, et elle est qualifiée en termes d'adaptation et d'institutionnalisation (Hakansson, 1982).

(b) **Les acteurs de l'échange**, appelés *participants* à l'interaction (le fournisseur et le client sont représentés avec leurs différentes caractéristiques susceptibles d'influencer le processus d'interaction). Comme leurs caractéristiques influencent la façon avec laquelle ils interagissent, Hakansson (1982) et Valla (1986) retiennent quatre dimensions d'analyse de ces participants : système social, organisation, stratégie, individus. La première prend en compte l'entreprise comme groupe social, la deuxième « organisation » englobe trois dimensions que sont la technologie, la structure et l'organisation de l'entreprise, la troisième sa stratégie, et la dernière « individus » concerne les caractéristiques des individus en présence : statut, fonction, expérience, objectifs, attentes.

(c) **L'atmosphère** de la relation renvoie au climat qui règne entre les deux organisations ; ce concept est utilisé "*pour qualifier l'environnement direct de l'interaction dans ses dimensions psycho-socio-culturelles*" (Valla, 1987). L'atmosphère de la relation peut être caractérisée en termes de pouvoir et de dépendance entre acteurs de la relation, de niveau de conflit et de coopération, ou d'opportunisme et de distances sociales et de niveau de compréhension entre les organisations (Hallen et Sandström, 1988). Complétant ces définitions, Michel et al. (1996) définissent l'atmosphère de la relation comme « le climat qui s'instaure entre les parties. Elle est constituée de

dépendance réciproque ou du sentiment de cette dépendance, d'orientation coopérative ou conflictuelle, de distance culturelle et sociale, de confiance accordée ou non ».

(d) **L'environnement** de l'interaction (regroupant les éléments autres que ceux liés directement aux deux organisations en présence, comme par exemple des éléments du contexte politique, culturel, social, de la structure du marché, de son dynamisme...) : il peut être décrit en termes de réseau, et on peut y ajouter des dimensions d'analyse telles que le contexte politique et économique, culturel et social, la structure des marchés, leur degré d'internationalisation, leur dynamisme (Hakansson, 1982 ; Salle et Sylvestre, 1992). La relation fournisseur/client est influencée par les caractéristiques du réseau dans lequel elle est insérée, i.e. par le type d'acteurs présents, la densité des liens entre eux, les règles du jeu en vigueur, etc. Elle résulte aussi des stratégies des deux entreprises en présence pour établir, maintenir ou faire évoluer leur position respective au sein de ce réseau d'acteurs (figure 3) :

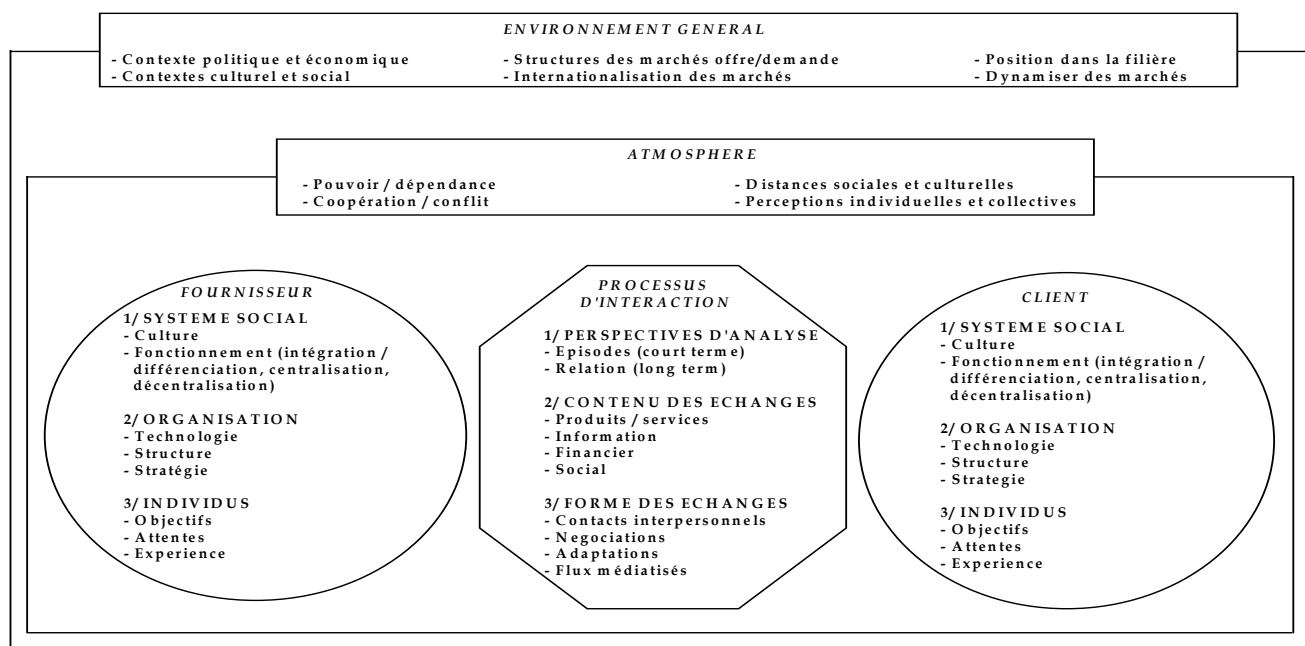


Figure 3 : Le modèle d'interaction. (Sources : Hakansson, 1982 et Valla, 1987)

### 3.2.4 L'innovation comme produit de l'interaction client/fournisseur :

Si les recherches menées au sein du groupe IMP, ne sont pas focalisées sur l'innovation, un certain nombre de travaux a tout de même été réalisé dans ce champ dans les années 1980, en ce qui concerne l'industrie (Hakansson, 1987), ou par des chercheurs proches de ce groupe (Millier,

2000). Ensuite, à partir des années 2000, des travaux ont été conduits par certains membres du groupe IMP, principalement dans le secteur des technologies de l'information (Blomqvist, Tikkanen, Möller, 2002 ; Hakansson et Waluszewski, 2007).

Hakansson (1987) analyse différentes formes de développement technologique de produit. Après avoir évoqué le développement de produit traditionnel initié par le producteur (paradigme « manufacturer active »), il note à propos du développement de produit initié par l'utilisateur que ce phénomène a été principalement mis en évidence par von Hippel (1976, 1986), et qu'il a aussi été mis en évidence dans le BtoB par Axelsson & Hakansson(1984). Hakansson (1987) s'appuie sur l'existence d'innovations conçues par les utilisateurs dont traitent ces travaux, pour remettre en cause le modèle traditionnel d'innovation par le fournisseur, et pour proposer ensuite une approche de l'innovation selon laquelle celle-ci serait le résultat d'un processus d'interaction entre client et fournisseur, tous deux prenant une part active aux processus d'innovation. Ainsi dans une étude réalisée par les chercheurs du groupe IMP en milieu industriel (Hakansson, 1987) des processus de développement conjoints purent être identifiées dans la majorité des cas.

Les trois types de *product development*, que nous venons d'évoquer, sont complémentaires, mais selon certaines critiques ne couvrent pas tout le spectre du développement produit. Une critique concerne le fait que la recherche en *product development* se focalise sur un ou quelques acteurs, seulement. Cette critique est formulée notamment par les tenants d'une approche « réseau » qui affirment que l'entreprise doit être considérée comme l'élément d'un réseau, et non comme une unité indépendante (Cook et Emerson, 1978). Reprenant cette approche, certains chercheurs du groupe IMP qui travaillent sur l'innovation (Hakansson, 1987), adoptent un point de vue sur ce thème, qui est d'abord une critique de l'approche de l'école diffusionniste (Rogers, 1962). Nous rappelons que cette école considère l'innovation comme préconstituée de façon autonome par le fournisseur, puis mise en relation avec son marché, déclenchant ainsi un processus d'adoption. Quelques éléments de l'approche de ces chercheurs sont regroupés sous le terme d'école «des réseaux d'innovation». Cette école est fondée sur l'insertion des entreprises dans des réseaux d'acteurs en interaction (entreprises et individus), et reprend la vision développée par certains sociologues de l'innovation, dans le cadre notamment de l'*Actor-Network Theory* ou ANT (Callon, 1981 ; Callon et Latour, 1985 ; Callon, 1998). Pour mener à bien leur activité, les entreprises cherchent à la fois à accéder à des ressources et à en développer d'autres en s'appuyant sur les relations qui existent entre elles ou en en créant de nouvelles. L'innovation suit ce type de fonctionnement : elle s'insère et se construit au sein de ces réseaux d'acteurs liés entre eux par des

positions relatives, des processus de travail, des échanges sociaux, informationnels et techniques que l'innovation va contribuer à bouleverser. Pour mener à bien un projet d'innovation, l'entreprise va donc associer un ensemble d'acteurs qui y trouvent un intérêt. Ils vont se l'approprier et lui donner forme, « la traduire » (Latour, 1989). Les réseaux ont des structures préexistantes de densité variable selon la dépendance des acteurs entre eux. En conséquence, l'innovation va être confrontée, même si son degré de différenciation technique par rapport à l'existant est fort, à des facteurs de réussite d'une autre nature liés à l'intérêt que les acteurs peuvent avoir de se l'approprier. En effet, l'innovation peut nécessiter la constitution d'un nouveau réseau impliquant à la fois des acteurs déjà présents et d'autres, et modifier de façon drastique la structure des réseaux. Dans les projets de R&D de forme « *market pull* », généralement issus de l'analyse des besoins des clients, l'innovation change en général peu la structure des réseaux d'acteurs présents. Par contre, les projets de forme « *technology push* », généralement issus de travaux de recherche menés par les entreprises, s'accompagnent souvent de la création d'un nouveau réseau, qui s'impose aux réseaux existants (Hakansson, 1987).

Pour Hakansson (1987), les réseaux jouent un rôle primordial quant au développement d'innovation technologique. Il définit les réseaux comme la toile de contacts qui existent entre les fournisseurs, les clients et les producteurs, en fait, l'ensemble des acteurs (individus, organisations, ...) d'une industrie. L'auteur considère le développement technologique comme le résultat de l'interaction entre différentes entreprises, organismes et individus, plutôt que comme la performance d'un acteur unique, en écho au concept de « *Market as Network* » (Johansson et Mattsson, 1987). Le développement technologique est alors considéré comme partie intégrante de la structure de laquelle il émerge. Les marchés industriels étudiés par les chercheurs du groupe IMP, se caractérisent par un degré élevé de division du travail sous forme d'activité industrielle menée par les entreprises (Hagg et Johanson, 1983). Ces activités sont plus ou moins dépendantes les unes des autres, et constituent des chaînes ou des filières industrielles. De ce fait, la performance d'une entreprise dans une activité industrielle donnée est conditionnée par la performance des autres entreprises exerçant des activités industrielles en amont et en aval. Pour mener à bien ces activités industrielles, les entreprises doivent donc disposer de ressources internes ou se les procurer dans ce gigantesque réseau constitué par l'ensemble des organisations présentes au sein des marchés industriels (Johanson, 1989). Selon les choix stratégiques d'une entreprise, l'accès à ces ressources se fera via les activités productrices au sein de l'organisation ou via la relation avec d'autres organisations. De ce fait, les entreprises vont être plus ou moins dépendantes les unes des autres (Hakansson, 1987).

Ainsi, pour Hakansson (1987), l'interaction au sein des relations fournisseur/client est critique pour le développement des connaissances, la mobilisation des ressources, et la coordination des ressources. A l'image du développement technologique vu comme un problème interne, Hakansson (1987) oppose l'idée qu'une grande part du processus de développement s'effectue sous la forme d'échanges techniques entre acteurs, individus et entreprises. L'innovation ne devrait alors pas être vue comme le produit d'un seul acteur, mais comme le résultat d'un jeu entre deux acteurs ou plus, comme le produit d'un réseau d'acteurs. Cette conclusion se fonde sur trois arguments principaux liés au développement de la connaissance, à la mobilisation des ressources, et la coordination des ressources (Hakansson, 1987). Nous détaillons ci-après ces trois arguments :

**1. Le développement des connaissances :** la connaissance nouvelle (concernant les produits nouveaux) émerge souvent à l'interface entre deux aires de connaissances différentes. Dans des situations d'échange, différents types de connaissances sont réunis (combinées ou confrontées) pour créer des situations innovantes. Le cas où l'échange a lieu entre l'acheteur et le vendeur constitue un cas particulier: Les besoins du client sont confrontés aux solutions techniques potentielles du vendeur, ce qui permet de réviser à la fois les besoins et les solutions, et de cette manière de trouver de nouvelles possibilités. Cependant, toutes les situations ne produisent pas obligatoirement des innovations.

**2. La mobilisation des ressources :** dès qu'une invention se matérialise sous la forme d'un produit, système ou service, elle devient dépendante d'autres produits, systèmes ou services. De telles dépendances existent entre l'invention et les acteurs qui doivent apprendre à l'utiliser, et à les combiner avec d'autres produits. Cela signifie que les acteurs doivent s'adapter pour apprendre à utiliser l'invention et en faire une innovation. Cela signifie aussi que l'innovation doit être adaptée, révisée et reconçue afin de devenir utile dans différentes situations, Le processus d'innovation comporte ainsi des éléments d'apprentissage, d'adaptation et de socialisation. Tous ces éléments demandent des ressources, dans un environnement où elles sont généralement rares, et où d'autres activités sont en compétition pour les mêmes ressources. En conclusion, selon Hakansson (1987), le processus d'innovation peut être comparé à un processus de mobilisation, qui à son tour doit être basé sur des processus d'interaction extensifs.

**3. La coordination des ressources :** alors que la masse des connaissances croît régulièrement, les ressources des entreprises restent limitées, ce qui les conduit à spécialiser leurs ressources de

développement dans des domaines restreints : inutile d'essayer de développer des produits seuls, alors qu'ils peuvent être développés plus efficacement ailleurs. Ce processus de spécialisation entraîne la spécialisation des moyens de production : il en résulte un grand nombre de petites unités de développement spécialisées, qui demandent à être coordonnées. Cette coordination doit être gérée au travers une série de relations d'échange entre ces unités. La coopération est aussi une façon de mobiliser des ressources externes (par exemple, chez un « gros » client engagé dans le processus d'innovation, qui peut ensuite faire des tests). Ces différents éléments peuvent parfois conduire une entreprise à considérer la coopération technique comme la seule possibilité de développement technologique. Cependant la coopération peut aussi créer des problèmes quant il s'agit de gérer un processus de développement à plusieurs acteurs. Mais le problème le plus important est politique, car l'entreprise doit gérer des relations englobant à la fois des éléments de coopération et de conflit.

On pourrait affirmer que cette conception du processus d'innovation s'oppose diamétralement à celle de la littérature sur la modularité, qui souligne l'indépendance des acteurs dans le processus d'innovation modulaire, la coordination étant alors « encadrée » (Sanchez et Mahoney, 1996), et les différents responsables de module, libres d'innover indépendamment des autres. Toutefois, comme notre travail de recherche s'intéresse à un stade du processus d'innovation où l'architecture modulaire de l'innovation n'est pas complètement définie, nous considérons que cet encastrement n'est pas encore réalisé, et que des interactions entre acteurs de ce processus seront nécessaires à ce processus.

### **3.3 Conclusion :**

Les travaux dans le champ du marketing interorganisationnel, se focalisent sur l'analyse des relations entre acteurs organisationnels, plutôt que sur celle des transactions, pour étudier les déterminants de ces dernières. Chaque relation fournisseur/client englobe les multiples d'interactions entre acteurs situés chez le fournisseur et chez le client. L'étude de la relation fournisseur/client, requiert donc l'étude de ces différentes interactions, chacune d'entre elles pouvant être caractérisée grâce au modèle d'interaction, outil élaboré par les travaux des chercheurs du groupe IMP (Hakansson 1982 ; Valla, 1987), que nous nous proposons d'utiliser pour caractériser les interactions fournisseur/client, mais aussi les interactions entre fournisseurs, qui sont à l'œuvre dans le processus d'innovation qui nous intéresse.

Nous tenterons de comprendre comment la perspective adoptée par les chercheurs du groupe IMP, qui considère l'innovation comme le produit d'un réseau d'acteurs en interaction, peut se conjuguer avec la perspective adoptée dans le champ de l'innovation par les utilisateurs, qui attribue à certains utilisateurs présentant des caractéristiques de « *lead users* », un rôle central dans le processus d'innovation *lead user*. Nous chercherons à savoir ensuite dans quelle mesure la première perspective est susceptible d'enrichir la seconde, cette dernière s'en distinguant, notamment, par le caractère multidimensionnel qu'elle attribue à la relation entre un client et un fournisseur.

#### 4. CONCLUSION SUR LES APPORTS DE LA LITTÉRATURE :

En conclusion de ce chapitre, nous insistons sur les principaux questionnements qu'a fait émerger l'analyse des littératures relatives aux champs que nous avons retenus : ceux de la modularité, de l'innovation par les utilisateurs, et du marketing interorganisationnel.

La littérature sur la modularité, insiste sur le fait que le logiciel est un des champs d'application, si ce n'est le champ d'application idéal pour ce concept. Elle décrit des architectures modulaires dont la définition passe par la définition de *design rules* ou règles de conception. Toutefois, le processus qui conduit à la définition précise de ces *design rules* a très peu été traité par cette littérature (Baldwin et Clark, 2000). Nous chercherons donc à caractériser le processus qui conduit à leur émergence, à cerner quels en sont les acteurs, et à déterminer les modalités selon lesquelles ils agissent.

Le logiciel constitue aussi, selon la littérature, un terrain d'application privilégié du concept d'innovation par les utilisateurs (Franke et von Hippel, 2003). Le logiciel est indubitablement un terrain favorable à l'innovation utilisateur, de par son architecture modulaire et de par son immatérialité, toutes choses qui facilitent les contributions d'un grand nombre d'acteurs, éventuellement indépendants, au processus d'innovation. Dans ces conditions les auteurs de ce courant de la littérature, affirment que ce sont les utilisateurs, présentant les caractéristiques de *lead user*, et non les éditeurs de logiciel, qui sont les mieux placés pour innover, en tant qu'expert de leurs propres usages et besoins. Mais, dans le cas des innovations logicielles de rupture, fondées sur des principes technologiques nouveaux, il se pose certaines questions que cette littérature aborde peu. Ainsi, par exemple, comment les *lead users* peuvent-ils surmonter les difficultés liées à l'apprentissage de ces principes technologiques ?

Le processus de conception d'un logiciel innovant dont chaque composant a été conçu antérieurement au produit par des organisations indépendantes, correspond à un processus typique de l'industrie logicielle, où l'innovation est portée par de petites structures spécialisées (Cusumano, 2004). Ce processus d'innovation logicielle radicale multi-acteur génère de nombreuses questions. Tout d'abord, qui joue le rôle d'architecte de la solution ? Un éditeur de logiciel, un client, ou encore une société de services informatiques ? Ensuite, la conception et la mise en œuvre de ce type d'innovation nécessite ce que nous appelons une double intégration : l'intégration des différents logiciels entre eux pour constituer un produit nouveau, et l'intégration de ce produit au système d'information des clients. Comment se déroule ce double processus d'intégration ? Quels en sont les acteurs ? Quelles sont leurs compétences, comment se constituent-elles ? Ce processus doit permettre la finalisation des *design rules* de l'innovation, comment celles-ci émergent-elles ? Dans un processus où la nouvelle architecture associe des acteurs qui ne coopéraient pas auparavant, des interactions vont émerger, comment se structurent-elles et comment se mettent-elles en place ?

Cela nous amène à prendre en compte les recherches effectuées dans le champ du marketing interorganisationnel, qui se sont focalisées sur l'analyse des relations entre acteurs organisationnels. Chaque relation fournisseur/client englobe de multiples d'interactions entre acteurs, interactions caractérisées grâce à un outil, le « modèle d'interaction », élaboré par des chercheurs du groupe IMP (Hakansson 1982). Nous nous proposons d'utiliser ce modèle pour caractériser les interactions à l'œuvre dans le processus d'innovation qui nous intéresse. Cela doit nous permettre d'affiner notre compréhension du processus d'innovation, et en particulier des rôles joués par les différents acteurs. Nous tenterons de comprendre comment la perspective adoptée, entre autres, par certains chercheurs du groupe IMP, qui considèrent l'innovation comme le produit d'un réseau d'acteurs en interaction, peut se conjuguer avec la perspective adoptée par les tenants de l'innovation par les utilisateurs qui attribue aux *lead users*, un rôle crucial dans le processus d'innovation *lead user*. Nous chercherons à savoir dans quelle mesure la première perspective, qui prend notamment en compte la multiplicité des acteurs ou groupes d'acteurs qui participent à la relation entre une entreprise et son fournisseur, est susceptible d'enrichir la perspective *lead user*, perspective différente de la première en ce qu'elle décrit une relation quasi-monodimensionnelle entre un fournisseur et un *lead user* (individu ou organisation).

Les questionnements retenus ont conditionné les choix méthodologiques que nous abordons au chapitre suivant, sachant que les opportunités d'accès au terrain qui nous ont été offertes, n'ont

pas été totalement indifférentes à l'émergence de ces questionnements. Ces choix méthodologiques nous ont conduit à réaliser une étude empirique, démarche peu fréquente dans le champ de la modularité (Fixson, 2003).

# **Chapitre 3 : Méthodologie**

## SOMMAIRE DU CHAPITRE 3 :

<b>Chapitre 3 : Méthodologie</b>	<b>101</b>
<i>Sommaire du Chapitre 3 :</i>	<i>102</i>
1. <i>Les démarches de recherche en gestion et processus d'innovation :</i>	<i>103</i>
1.1 Introduction :	103
1.2 Définition d'un cadre épistémologique et intégrateur :	103
1.3 Les méthodes mises en œuvre dans la recherche sur la conception :	104
1.4 Une approche processuelle :	107
1.5 Conclusion :	108
2. <i>Le questionnement de l'entreprise :</i>	<i>108</i>
2.1 Projet de recherche :	108
2.2 Questionnement de l'entreprise :	109
2.3. Construction de la place du chercheur sur le terrain :	111
3. <i>Les choix méthodologiques :</i>	<i>115</i>
3.1 Une innovation développée au travers de deux projets :	116
3.2 Le projet comme unité d'analyse :	116
3.3 Une analyse processuelle :	117
3.4 L'étude d'une situation emblématique de l'industrie du logiciel :	119
3.4.1 L'étude de la conception d'un nouveau type de logiciel :	119
3.4.2 Les logiciels de « <i>text mining</i> », innovation radicale (ou innovation d'exploration) :	120
3.4.3. L'accès à une situation emblématique : innovation modulaire multi-acteurs.	121
3.5. Modalités de l'étude – recueil de données :	122
3.5.1 Modalités de l'étude :	122
3.5.2 Les données recueillies :	124
3.5.3 Les limites de la mise en œuvre de la méthodologie retenue :	128
3.5.4 Conclusion de la partie « Modalités de l'étude - Recueil de données » :	130
4. <i>Conclusion : une opposition observation/intervention à relativiser.</i>	<i>130</i>

Nous exposons maintenant la méthodologie que nous avons employée, en décrivant dans un premier temps la démarche méthodologique retenue, puis en présentant les entreprises étudiées et les données collectées.

## **1. LES DEMARCHES DE RECHERCHE EN GESTION ET PROCESSUS D'INNOVATION :**

### **1.1 Introduction :**

Nos travaux s'inscrivent dans une tradition méthodologique de recherche intervention et d'observation participante, dans laquelle le chercheur privilégie la relation longue et étendue avec l'entreprise, à l'étude d'un échantillon très important, typique des méthodes quantitatives largement répandues en sciences de gestion. Comme annoncé en introduction, nous avons mené une analyse portant sur une innovation radicale dans le secteur des logiciels destinés aux entreprises. Une méthode de recherche peut s'appréhender comme une série de choix permettant d'opérationnaliser une stratégie de recherche. Cette stratégie de recherche pouvant être décrite, dans le sens de Yin (2003) comme une « démarche structurée dont l'issue est une représentation et une explication d'un phénomène complexe » (Wacheux, 1996).

### **1.2 Définition d'un cadre épistémologique et intégrateur :**

Avant de présenter en détail la démarche que nous avons adoptée en fonction de notre objet de recherche, nous décrivons des méthodes qualitatives et insistons sur ce type de démarche pour étudier les processus d'innovation.

David, Hatchuel et Laufer (2001) distinguent quatre démarches de recherche en sciences de gestion l'observation participante (ou non), la recherche-action, la conception « en chambre » de modèles et outils de gestion, et la recherche-intervention Ils placent ces quatre types dans une matrice en distinguant en colonne l'objectif de construction mentale et l'objectif de construction concrète ; et en ligne, une démarche qui partirait de l'existant (observation des faits ou travail d'un

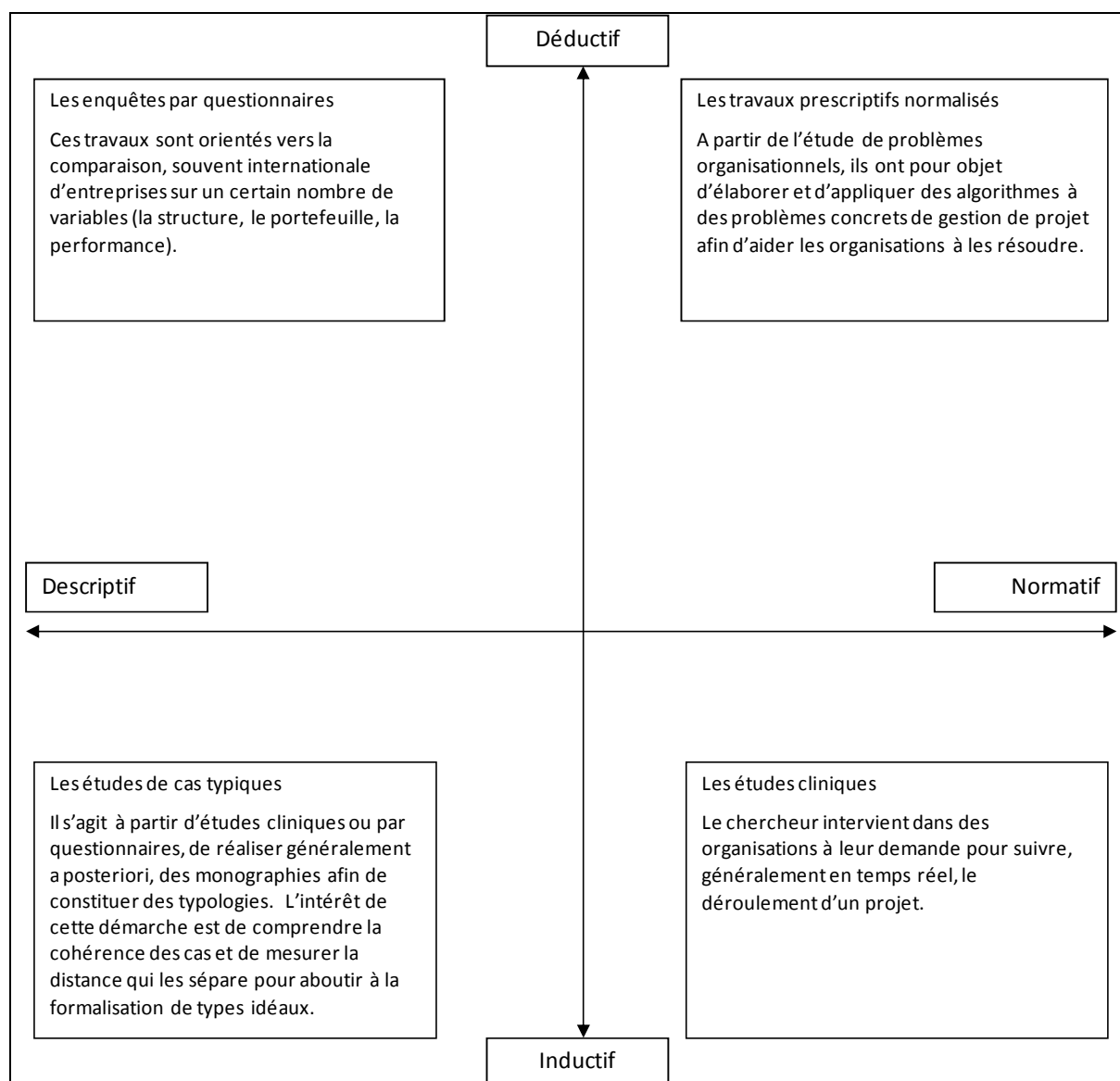
groupe sur lui-même) et une démarche qui partirait d'un objet de transformation au moins partiellement défini (figure 4) :

		Objectif	
		<i>Construction mentale de la réalité</i>	<i>Construction concrète de la réalité</i>
<b>Démarche</b>	<i>Partir de l'existant (observation des faits ou travail du groupe sur son propre comportement)</i>	<b>Observation, participante ou non (I)</b> Élaborer un modèle descriptif du fonctionnement du système étudié.	<b>Recherche-action (II b)</b> Aider à transformer le système à partir de sa propre réflexion sur lui-même, dans une optique participative.
	<i>Partir d'une situation idéalisée ou d'un projet concret de transformation</i>	<b>Conception « en chambre » de modèles et outils de gestion (II a)</b> Élaborer des outils de gestion potentiels, des modèles possibles de fonctionnement, sans lien direct avec le terrain.	<b>Recherche-intervention (III)</b> Aider, sur le terrain, à concevoir et à mettre en place des modèles et outils de gestion adéquats, à partir d'un projet de transformation plus ou moins complètement défini.

Figure 4 : un cadre intégrateur pour quatre démarches de recherche en sciences de gestion. (David, A., in David, A., Hatchuel, A. et Laufer, R, 2001).

### 1.3 Les méthodes mises en œuvre dans la recherche sur la conception :

Lenfle (2001, p13) souligne quant à lui la diversité des méthodes de recherche comme étant caractéristique des recherches sur la conception. Ainsi, Clark et Fujimoto (1991) combinent enquête par questionnaires et entretiens directs avec les acteurs, tandis que les recherches sur l'industrie informatique recourent à l'étude comparative de différents cas (Brown & Eisenhardt, 1997) et que la majorité des études françaises se fondent sur des méthodes de recherches cliniques (Midler, 1993 ; Garel, 1994 ; Weil, 1999). Cette coexistence de logiques inductive et déductive à visées tantôt normatives tantôt descriptives a été schématisée par Garel (1998) de la manière suivante (figure 5) :



**Figure 5 : Méthodes mises en œuvre dans la recherche sur la conception** (Garel, 1998).

La pertinence d'une méthode de recherche se jugeant par son adéquation à l'objet étudié, il est logique d'admettre un « pluralisme méthodologique contrôlé » (aucune méthode n'a a priori, le monopole de la rigueur et de la raison) qui veille au travail conceptuel, à l'explication de ses présupposés, à la pertinence, à la cohérence et à l'efficacité des modélisations, à la lisibilité des cheminements entre termes théoriques et termes empiriques, à la communication des énoncés... » (Martinet, 1990, p.23). C'est cette exigence de contrôle et de lisibilité qui nous amène à expliquer les fondements de notre méthode de recherche et à exposer son déroulement.

Benghozi, Charue-Duboc et Midler (2000) soulignent l'intérêt de méthodologies spécifiques qui permettent de prendre en compte un certain nombre de caractéristiques des processus

d'innovation qui les conduisent à mettre en évidence l'intérêt de méthodologies spécifiques. Les principaux enseignements retirés par les auteurs sont les suivants :

- **Des analyses rétrospectives sur une longue période** : l'idéologie de la « nouveauté » fait partie de la rhétorique de l'innovation. Il est alors important de resituer les transitions étudiées dans les approches longitudinales des évolutions à plus long terme (Paris, 2002 ; Ben Mahmoud-Jouini & Midler, 1999).

- **Une confrontation de situations emblématiques et différenciées** : le choix de la situation de conception à étudier pose question. La question de la contingence (ou au contraire de la généralité) des modèles théoriques élaborés constitue l'une des questions récurrentes de la discipline de gestion. L'approche adoptée peut être rapprochée de l'usage de la notion de « fait stylisé » tel qu'on l'entend en économie. Les situations étudiées ne sont pas choisies comme des représentants moyens ou « représentatifs » d'une population. On cherchera au contraire des situations « emblématiques » où le problème abordé revêt un caractère extrême, où les acteurs se proposent d'expérimenter des démarches ou des organisations qui, justement, n'existent pas ailleurs. La qualité du modèle (fait stylisé) s'évaluera à sa cohérence interne (testée empiriquement dans la situation analysée) et à sa mise en relation avec une caractérisation précise du contexte. Cet accent mis sur la situation est à rapprocher du courant constructiviste, pour qui *« les problèmes ne se posent pas d'eux-mêmes. C'est précisément ce sens du problème qui donne la marque du véritable esprit scientifique. Pour un esprit scientifique, toute connaissance est une réponse à une question. S'il n'y a pas de question, il ne peut pas y avoir de connaissances scientifiques. « Rien ne va de soi, rien n'est donné. Tout est construit »* (Bachelard 1970). La confrontation d'une variété suffisante de situations est alors la condition nécessaire pour absorber les différentes dimensions du phénomène étudié.

- A propos de la question de la **généralisation** à partir de l'étude d'une situation emblématique, Charue-Duboc (2007) considère que la démarche du chercheur « consiste à identifier des principes qui sont sous-jacents aux solutions contingentes mises en œuvre et à construire par l'analyse une relation entre le principe organisationnel de fonctionnement et le sous-problème caractérisé [...] à élaborer des énoncés, modèles, catégories analytiques qui peuvent être mobilisés dans d'autres situations que celles qui ont permis de les construire. Cela suppose d'avoir cherché à caractériser la situation et d'avoir formulé les résultats d'une manière qui puisse être transposée à d'autres situations. »

- **Des recherches interactives sur de longues durées avec les firmes :** l'innovation est une logique qui s'étudie mal a posteriori. Il faut être confronté à l'incertitude des débats, à la pluralité des possibles ex ante pour comprendre l'intérêt et les limites des dispositifs ou instrumentations qui ambitionnent d'améliorer les capacités de conception collective. D'où le recours à des démarches d'accompagnement en temps réel des projets. Par ailleurs, l'analyse des transformations des systèmes de conception implique une interaction sur des échelles de temps compatibles avec la durée des changements étudiés. Dès lors, les collaborations se déploient sur quatre ou cinq ans, souvent plus. L'interaction longue avec la situation étudiée est probablement le seul moyen pour « avoir une chance d'obtenir des échantillons prélevés au hasard, de disposer d'un éventail d'événements inattendus et d'arriver à un degré de familiarité étroite avec les gens » (Goffman 1974) ». Cette inscription sur la durée est également soulignée par Chanal, Lesca et Martinet (1997) « la dimension temporelle et progressive apparaît comme un point central de cette démarche qui soumet les acteurs à des processus d'assimilation-accommodation leur permettant une modification de leur représentations antérieures ».

#### 1.4 Une approche processuelle :

Une théorie du processus (Van de Ven et Poole, 1989) a émergé qui traite de méthodologies appropriées aux processus d'innovation, qui est à rapprocher des travaux de Benghozi, Charue-Duboc et Midler (2000). En effet, les travaux de Van de Ven et Poole (1989) prônent la réalisation d'analyses processuelles, recommandation qu'ils justifient en mettant en avant des éléments qui peuvent être rapprochés de ceux qui sont soulignés par Benghozi, Charue-Duboc et Midler (2000). Van de Ven et Poole (1989) soulignent que si de nombreuses études ont examiné les antécédents ou conséquences de l'innovation, très peu ont directement examiné comment et pourquoi les innovations émergent, se développent, croissent, et s'achèvent avec le temps. Une appréciation d'une séquence temporelle des activités dans le développement et l'implémentation de nouvelles idées est fondamentale au management de l'innovation. Ils soulignent que les managers de l'innovation ont besoin de connaître plus que les facteurs expliquant les succès. Ils ont besoin d'une « *road map* », qui indique comment et pourquoi la trajectoire d'innovation se déroule, et quels chemins ont conduit au succès ou à l'échec. En d'autres termes, le manager de l'innovation a besoin d'une théorie du processus qui explique le développement de l'innovation.

C'est une théorie du processus qui se focalise sur l'explication de l'ordre temporel et, de la séquence d'étapes qui permet de transformer et de mettre en œuvre une idée innovante en une

réalité concrète. Pour bâtir une telle théorie, les auteurs conseillent d'étudier l'ensemble du processus d'innovation et de retenir les cinq catégories d'observations suivantes : les idées, les personnes, les transactions (ou interactions), le contexte et les résultats (Van de Ven, 1986). L'auteur justifie ces catégories par le fait que le processus d'innovation consiste en la motivation et la coordination de personnes qui développent et mettent en œuvre de nouvelles idées en engageant des transactions (ou des relations), avec d'autres, et en faisant les adaptations requises pour atteindre les résultats désirés à l'intérieur de contextes institutionnels et organisationnels changeants (Van de Ven, Angle et Poole, 1989). En soulignant l'importance de recherches sur les processus, ils rejoignent les recommandations méthodologiques que nous venons de rappeler, et notamment la troisième, qui concerne les recherches interactives sur des durées longues avec les firmes (Benghozi, Charue-Duboc et Midler, 2000).

## **1.5 Conclusion :**

La prise en compte de différents travaux relatifs aux méthodologies adaptées à l'étude des processus d'innovation, nous conduit à réaliser une analyse processuelle, grâce à une interaction sur une longue période avec l'entreprise.

## **2. LE QUESTIONNEMENT DE L'ENTREPRISE :**

Nous présentons maintenant notre projet de recherche, le questionnement de l'entreprise, et la construction en plusieurs phases de notre place de chercheur.

### **2.1 Projet de recherche :**

Notre question de recherche est la suivante « quels rôles jouent les premiers clients dans le processus de conception d'innovation logicielle radicale multi-acteur, et comment les associer à ce processus ? ». Les champs que nous mobilisons pour traiter cette question sont ceux de la modularité, de l'innovation par les utilisateurs, et du marketing interorganisationnel.

Notre question de recherche est la suivante « quels rôles jouent les premiers clients dans le processus de conception d'innovation logicielle radicale multi-acteur, et comment les associer à ce processus ? ». Les champs que nous mobilisons pour traiter cette question sont ceux de la modularité, de l'innovation par les utilisateurs, et du marketing interorganisationnel.

Notre projet de recherche vise donc à décrire et comprendre les logiques selon lesquelles la collaboration fournisseur/client et les collaborations entre fournisseurs se mettent en place, et comment le client participe au développement final du projet innovant. La méthode de recherche que nous présentons nous a paru adaptée à la fois au contexte du projet de recherche (mariage des deux champs de littérature, l'un relatif à l'innovation par les utilisateurs et l'autre à celui de la modularité, en intégrant les apports du marketing interorganisationnel), au thème de la recherche et aux objectifs de la recherche (repérer l'originalité des collaborations et en identifier les implications sur la gestion de projet par le fournisseur et l'adoption de l'innovation par le client).

## 2.2 Questionnement de l'entreprise :

A l'origine de cette question de recherche, au début des années 2000, on trouve le questionnement de l'entreprise Xerox, plus précisément des dirigeants du centre de recherche XRCE situé près de Grenoble, centre spécialisé dans le traitement du langage ou *language processing*. Notre relation avec le XRCE (Centre de Recherche Européen de Xerox) date de 2001, année où nous avons aidé les responsables marketing de deux projets de *text mining* (ou *text analytics*, i.e. analyse automatisée de texte) développés au XRCE à analyser leur marché potentiel, et plus précisément à le segmenter. Des questionnements des managers des équipes-projet, devenues « *business units* », (« *manager* », « *market manager* », « *sales manager* », « *technical manager* ») ont alors clairement émergé : ils concernaient la difficulté de leur client à comprendre, adopter et à utiliser à bon escient les logiciels innovants, mais aussi la difficulté rencontrée par les *business units* à prendre en compte les attentes et les besoins des clients pour développer des offres personnalisées et à proposer de nouveaux logiciels. Il y avait donc au départ une insatisfaction « marketing » (« comment comprendre et convaincre le client ? »), et une insatisfaction liée à la création d'offres nouvelles (« comment orienter et optimiser le développement d'une offre innovante ? »). Il y avait aussi une insatisfaction liée au fait que les business unit concernées travaillaient dans l'urgence : urgence de trouver des clients, de leur expliquer l'innovation (le terme « évangelisation » étant généralement utilisé par les professionnels de ce secteur) et les bénéfices qu'ils pourraient en retirer, mais aussi l'urgence de réaliser le développement d'une offre personnalisée adaptée à des besoins rarement explicites. Cette difficulté à s'extraire du court-terme limitait la prise de recul nécessaire à la définition de choix stratégiques relatifs à l'évolution de l'offre, mais aussi à la *business unit* en général. De ces insatisfactions a émergé un intérêt pour un travail d'étude visant globalement à mieux comprendre les clients. Grâce à cette première phase de contacts, avec les acteurs de différents projets de *text mining* développés au

XRCE, sur laquelle nous reviendrons, nous avons pu nous familiariser avec les acteurs (éditeurs de logiciels, centres de recherche publics ou privés, clients potentiels) et les technologies du *text mining*. Nous avons pu constater que les acteurs de l'industrie du logiciel travaillant dans des start-up innovantes, avaient la certitude que les premiers projets d'implémentation d'un logiciel présentant les caractéristiques d'une innovation radicale ou « innovation d'exploration », selon les auteurs (Chanal et Mothe (2005), Danneels (2002)) jouaient un rôle majeur dans le processus d'innovation. Selon eux, les premières implémentations permettaient la contextualisation d'un nouveau concept, et les utilisateurs jouaient un rôle unique quant à la transformation d'un nouveau concept d'application (le *text mining*) en un logiciel commercialisable (entretien avec Yves Mahé, Business Unit Manager, Xerox, janvier 2003 ; entretien avec Philippe Laval, directeur général de Sinequa, mars 2004), vision qui est aussi celle de la littérature (Cusumano, 2004). C'est donc cette question que nous avons cherché à explorer.

Le cas étudié se situe dans l'édition de logiciels professionnels destinés à la gestion électronique de document (GEIDE). Nous nous sommes intéressé au processus de conception de logiciels au sein de deux structures (« *Business Units* ») AskOnce et XTS/Xelda, créées pour le développement et la commercialisation de logiciels innovants conçus dans le cadre du XRCE. Ces deux structures ont essaimé en 2003 : l'une (XTS/Xelda) a été rachetée par Temis, éditeur français de logiciels de *text mining*, fondé par d'anciens cadres d'IBM en 2001, l'autre (AskOnce) a été rachetée par l'éditeur de logiciel américain Documentum, un des leaders du marché de la gestion de contenu (ou ECM, Enterprise Content Management).

Après quelques mois, les équipes essaimées au sein de nouvelles structures nous ont à nouveau accueilli, mais dans une perspective moins claire que précédemment. La demande de l'entreprise, en l'occurrence Temis, n'était alors plus aussi précise. Il s'agissait plus d'un intérêt global sur l'étude de l'implication des clients dans les projets les plus complexes. Au fur et à mesure de l'étude, un autre apport est apparu aux yeux de nos interlocuteurs, principalement chez Temis : l'intérêt de pouvoir faire le point sur les projets en cours, avec un interlocuteur « neutre », qui tente d'avoir une vision à la fois globale et structurée des projets. Un autre aspect a aussi facilité notre présence sur le terrain, que ce soit chez Temis, Mondeca ou leurs clients, le fait que nos interlocuteurs soient pour la plupart des chercheurs ou ex-chercheurs (en informatique, en sciences du langage, ou en gestion des connaissances) a créé une proximité et un intérêt a priori pour notre travail de recherche. Enfin, et à un degré moindre, le fait que notre travail conduise à des présentations publiques et à des publications, a été perçu comme un moyen de faire connaître

au « monde extérieur », l'innovation étudiée et les projets développés. Cet intérêt de Temis, puis de Mondeca s'est particulièrement manifesté pour les projets intégrant différentes briques logicielles de différents éditeurs de logiciels innovants.

L'intérêt de l'entreprise vient enfin de ce que ce type de projet nécessite une collaboration poussée avec d'autres éditeurs de logiciel. Or, les start-up logicielles focalisées sur une technologie innovante issue de la recherche travaillent dans un premier temps directement avec leurs clients en vue de sa mise en œuvre. Elles sont ensuite rapidement amenées à collaborer avec d'autres éditeurs de logiciels (innovants ou non), pour fournir des solutions globales, susceptibles de satisfaire leurs clients. Mais ces start-up logicielles n'ont, dans un premier temps, pas l'expérience de telles collaborations, et se posent des questions légitimes sur leur conduite, et sur la façon d'y intégrer le client. Il y a donc un questionnement à ce niveau, questionnement qui est renforcé par le fait que ces entreprises connaissent les logiques de plate-forme qui prévalent dans leur industrie. Elles savent donc qu'elles doivent s'efforcer de saisir les opportunités de participer à l'établissement de nouvelles plates-formes applicatives, problématique qu'elles n'ont pas non plus l'habitude de traiter. Cela éclaire donc leur intérêt pour une recherche portant sur la conception d'innovations couplées, susceptibles de devenir ou de faire partie d'un nouveau type de plate-forme logicielle.

L'intérêt de l'entreprise Temis, puis celui de Mondeca nous ont permis d'accéder au terrain, de nous entretenir de manière régulière avec un certain nombre d'acteurs clefs des projets étudiés, d'assister à diverses réunions en interne (notamment les réunions de lancement de projet), et d'assister à des présentations commerciales, mettant en valeur auprès de prospects des solutions emblématiques réalisées par Temis et Mondeca.

### **2.3. Construction de la place du chercheur sur le terrain :**

Nous revenons maintenant sur la façon avec laquelle notre position a été négociée. A ce propos, Girin (1990) rappelle trois raisons, spécifiques à la recherche en sciences de gestion, pour lesquelles « un degré élevé d'opportunisme » s'impose :

1. « D'abord, il faut entrer sur le terrain, et cette entrée se négocie. Cela peut se faire sur la base d'un projet de recherche élaboré au laboratoire, ou sur la base d'une demande d'intervention de la part d'un partenaire dans l'organisation, ou les deux, mais il y a toujours des compromis à faire.

2. Ensuite, il faut rester un minimum de temps. Pendant ce temps, il se passe des choses, ce qui est heureux pour notre instruction, mais souvent malheureux pour notre programme. Les interlocuteurs changent de place, de fonctions, de préoccupations, posent au chercheur de nouvelles questions, remettent en cause la possibilité de recherche. [...]

3. Enfin, la matière ne se contente pas de se mouvoir. Contrairement aux étoiles, elle pense. C'est très embêtant et très intéressant. Très embêtant, car la matière pense notamment à nous. Elle nous attribue des intentions qui ne sont peut-être pas les nôtres, mais qui va conditionner la manière dont elle va nous parler, ce qu'elle va choisir de nous montrer ou de nous cacher. Bref, la matière nous manipule, et risque de nous rouler dans la farine. Mais c'est très intéressant, car d'abord elle pense à notre profit, « pour nous ». On peut (on doit) être plus rigoureux, plus exigeant sur la forme et sur la cohérence de ce que l'on avance, que ne le sont les gens du terrain, mais on invente rarement un principe d'explication, une interprétation de leur situation et de leurs comportements, qui ne leur serait absolument pas venu à l'esprit. Bien au contraire, c'est un sentiment que j'ai souvent éprouvé, les gens de terrain sont des producteurs de théorie, des « savants ordinaires » auxquels il serait tout aussi stupide de ne pas prêter l'oreille, qu'il serait imprudent de prendre leurs raisonnements pour argent comptant. ».

Nous avons tenté de construire notre place de chercheur en faisant preuve d'opportunisme, tout en étant confronté aux difficultés et spécificités évoquées par Girin (1990). C'est ce que nous illustrons ci-après. Cette construction s'est réalisée au travers de quatre phases successives.

En effet, initialement en 2001, nous avons été contacté par Cyril Chantrier, en charge du marketing pour l'équipe-projet XTS/Xelda, afin d'analyser le marché potentiel des logiciels développés par cette équipe. Ce fut l'occasion de rentrer en contact et de travailler avec les marketers, commerciaux, chefs de projet, mais aussi pratiquement tous les membres de cette équipe-projet. Notre position était alors celle de consultant. Cette phase s'est déroulée entre mai 2001, et septembre 2002. Ensuite, nous avons négocié avec la direction du XRCE la possibilité d'effectuer un travail de recherche dans le cadre de notre thèse, en étudiant le processus d'innovation au travers de projets de mise en œuvre de logiciels innovants. En contrepartie implicite, nous devions apporter une expertise marketing et une vision extérieure favorisant le lancement commercial de ces produits. Dans cette deuxième phase, nous avons été amené à observer deux équipes-projet du XRCE : XTS/Xelda que nous avons déjà mentionné, et AskOnce. Nous avons ainsi pu participer à des réunions diverses (commerciales, marketing,

projet avec ou sans les clients concernés) organisées par ces *business units*. Mais mi-2003, les deux *business units* ont essaimé (rachat par mes éditeurs de logiciel Temis pour XTS/Xelda d'une part, Documentum pour AskOnce d'autre part). La direction du XRCE a changé, suite à la réorganisation de la recherche de Xerox en Europe, et la nouvelle direction n'a pas été intéressée par la poursuite de notre travail. Ce fut la même chose pour Documentum, éditeur de logiciel américain de grande taille (1000 employés). Nous avons ensuite pris contact avec la direction de Temis, qui a accepté après quelques mois, que nous reprenions notre travail de recherche en utilisant ses projets comme terrain : ce fut l'avant-dernière phase de construction de notre place de chercheur sur le terrain, elle a débuté en avril 2004. C'est lors de cette phase, que nous avons réellement commencé à collecter les données nécessaires à notre recherche. Lors de la dernière phase qui a débuté en octobre 2007, nous avons pu interroger différentes personnes de Mondeca : le président, la personne en charge du couplage entre les logiciels de Temis et de Mondeca, ainsi que le chef de projet, en charge du projet PressPro (achevé à ce moment-là) et Exinis (en cours, à cette époque). Nous avons par ailleurs pu continuer à interviewer le responsable du secteur « *publishing* » de Temis, ainsi que le responsable des partenariats OEM pendant cette phase.

**En synthèse, notre position sur le terrain s'est construite au travers de quatre phases successives :**

- **Phase 1 : mai 2001 - septembre 2002 :** travail sur la segmentation des marchés potentiels des logiciels utilisant les technologies de *text mining* au centre de recherche européen de Xerox (XRCE), en tant que consultant. Nous avons observé le fonctionnement des business units au travers d'interviews des responsables marketing et des managers, mais aussi travaillé directement avec les marketing managers des équipes-projet XTS/ Xelda et AskOnce, et organisé différentes réunions de travail portant sur la segmentation du marché, réunions en tant que consultant.
- **Phase 2 : octobre 2002 - juin 2003 :** participation à la vie de deux business units, AskOnce et XTS/Xelda, que Xerox désirait essaimer, en tant que chercheur en sciences de gestion. Dans cette phase, nous avons joué un rôle d'observateur, nous intégrant par moment complètement à la vie des équipes-projet, comme lorsque nous avons assisté aux réunions commerciales et marketing mensuelles du lundi matin, de la business unit AskOnce. Dans cette phase, Xerox attendait de nous, des avis et observations portant sur la façon de faire évoluer l'offre en tenant compte des marchés, en échange de cette

position proche de celle d'«étranger intégré» décrite par Girin (1990). Ces deux premières phases nous ont permis de nous familiariser avec le secteur du logiciel, les acteurs, les technologies et les marchés du *text mining*. Les informations recueillies en phases 1 et 2, nous ont donc permis d'acquérir une connaissance du contexte du secteur de la GEIDE et de ses acteurs, de nous familiariser avec la technologie du *text mining*, ainsi qu'avec différents responsables de Xerox.

Ces deux phases nous ont ainsi permis de créer des liens avec différentes personnes chez Xerox, qui ont facilité le passage à la troisième phase.

- **Phase 3 : avril 2004 - septembre 2007** : étude en tant que chercheur en sciences de gestion travaillant sur l'innovation auprès de Temis, après l'acquisition par cette start-up française, des droits d'exploitation des brevets de Xerox concernant les logiciels XTS (extraction terminologique) et Xelda (moteur d'analyse linguistique) et après le transfert l'équipe-projet en charge de son développement et de sa commercialisation de Xerox à Temis. Nous nous sommes alors focalisé sur la solution de « création automatisée de base de connaissance » et avons cherché à analyser le processus d'innovation, le projet, l'ensemble des différentes interactions avec le client de la manière la plus fine possible. Pour cela, nous avons interrogé régulièrement chez Temis les acteurs clefs en relation avec le client en question, dans le cadre d'entretiens semi-directifs. Il s'agissait en premier lieu du responsable des marchés « *publishing* » de Temis : c'est auprès de lui principalement, que nous avons pu, sur une base bimensuelle, collecter la plus grande partie de l'information nécessaire à notre recherche. Il s'agissait aussi du directeur technique « *core product* », et du « *OEM partnership manager* » responsable des partenariats OEM de Temis avec d'autres éditeurs de logiciel. Enfin, chez les clients, nous avons réalisé ponctuellement des entretiens semi-directifs en face à face, en posant des questions ouvertes aux personnes qui jouent ou ont joué un rôle quant à l'achat, la mise en œuvre (incluant une phase de développement personnalisé de la part du fournisseur), et l'utilisation des logiciels innovants dans leur entreprise. Il s'agissait généralement de représentants de la direction des systèmes d'information ou de la direction fonctionnelle concernée par l'innovation et d'utilisateurs de l'innovation.
- **Phase 4 : octobre 2007 - juin 2008**. Une situation identique à celle de la phase 3, avec de plus l'accès à différentes personnes chez Mondeca (président, responsable de couplage, chef de projet) que nous avons donc interviewées en tant que chercheur en

sciences de gestion, comme cela a continué d'être le cas chez Temis, auprès des mêmes acteurs qu'en phase 3. Il en a été de même, avec l'architecte innovation d'Exinis avec lequel nous avons pu nous entretenir pendant cette phase.

A l'origine, nous étions à la fois étranger à l'entreprise Xerox, mais aussi aux technologies du *document processing* en général, et du *text mining*, en particulier. Nous ne pouvions nous prévaloir que d'une courte expérience de chef de produit dans l'édition logicielle datant du début des années 1990. Nous notons toutefois que celle-ci nous a été utile dans la mesure où nous avons retrouvé chez Xerox, d'anciens collègues de cette époque. Toujours, est-il qu'il fallait avant d'entamer véritablement la collecte d'information relative à notre recherche, nous familiariser avec le secteur du *text mining*, ses acteurs, ses technologies, son vocabulaire. Cela n'a été possible que parce qu'au départ l'entreprise Xerox avait formulé une demande explicite de conseil de notre part. Le travail effectué dans la phase 1, puis dans la phase 2, a donc été précieux voire indispensable, pour préparer, notre travail de recherche, même si nous n'avions pas planifié les choses de cette manière.

Dans la même perspective, il faut parler des questionnements de Xerox, puis de Temis que nous avons déjà évoqués, l'interaction avec les acteurs clés des projets étudiés a été facilitée par un intérêt autre que celui d'apporter des éléments de réponse à ces questionnements. Nos interlocuteurs ont en effet perçu, très concrètement, un intérêt double à notre étude : l'un était de pouvoir faire un point sur l'avancement des projets avec un interlocuteur neutre, ce qui dans un environnement particulièrement turbulent, est rarement possible ; l'autre, plus diffus, était de contribuer à la communication autour d'une innovation complexe, mais techniquement et commercialement prometteuse pour Temis et Mondeca. Ces deux derniers points ont grandement facilité notre accès au terrain.

### **3. LES CHOIX METHODOLOGIQUES :**

Nous présentons maintenant l'innovation étudiée, ainsi que les projets qui ont permis sa conception et sa mise en œuvre. Nous expliquons ensuite les raisons qui justifient le choix du projet comme unité d'analyse, et revenons sur les raisons du choix d'une analyse processuelle. Enfin, nous expliquons pourquoi la situation étudiée, un processus d'innovation logicielle radicale multi-acteurs est emblématique de l'industrie du logiciel.

### 3.1 Une innovation développée au travers de deux projets :

Nous analysons le déploiement de cette innovation logicielle radicale (solution de création automatisée de base de connaissance) au travers de deux projets, au sein de deux grandes entreprises (PressPro, groupe de presse et Exinis France, éditeur). Chacun visait à développer une solution de création et de gestion de base de connaissance, basée sur des logiciels d'extraction, de catégorisation et d'archivage d'information. Nous avons pu suivre de manière régulière ces deux projets. Ces différents logiciels ont été développés principalement par Temis (qui propose les logiciels de *text mining* proprement dit) et une autre entreprise française de type start-up Mondeca (qui développe un système de gestion de base de connaissance). Il s'agissait dans ces deux projets de développer et d'implémenter une application de ces logiciels susceptible d'intéresser à terme tout groupe de presse et éditeur.

Le premier projet concerne la création d'une base de connaissance pour le groupe de presse PressPro. Il s'agissait de faciliter la création par les documentalistes du groupe, de dossiers de presse thématiques destinés aux journalistes. Il s'est déroulé de juin 2003 à mars 2006. Le deuxième projet porte sur la création d'une base de connaissance pour l'éditeur Exinis. Il avait pour but l'aide à la rédaction de notes de synthèses juridiques vendues aux professionnels du droit. Il a commencé en juin 2004, le déploiement de la solution a eu lieu fin 2007.

Ainsi nous sommes nous focalisé sur ces deux projets successifs visant à déployer des applications mobilisant une même brique logicielle innovante (elle-même composée de différents logiciels) chez différents clients. Nous comparons l'offre innovante et l'organisation de ces projets. Cette approche comparative vise à repérer des effets d'apprentissage et d'irréversibilité créés par les premiers produits innovants conçus pour les premiers clients par rapport aux produits de la même famille conçus pour les clients suivants.

### 3.2 Le projet comme unité d'analyse :

Le choix de notre unité d'analyse a été fait en relation avec notre question de recherche principale que l'on pourrait reformuler comme suit : « dans quelles circonstances et selon quelles modalités le client apporte-t-il une contribution déterminante au processus d'innovation radicale ? ». Etant donné cette question de recherche, nous avons décidé de choisir le projet comme notre unité d'analyse, en considérant que l'innovation « création automatisée d'une base de connaissance » va

émerger de projets successifs, en ce que ceux-ci vont permettre la définition d'une nouvelle offre produit (au travers de la personnalisation de briques logicielles existantes, de l'interfaçage entre celles-ci) et l'apprentissage relatif à l'ingénierie et à l'intégration de la solution chez le client :

- Une des caractéristiques des projets quand on les compare à d'autres formes d'échange, est leur temporalité. Le choix du projet comme unité d'analyse aide à définir les limites du processus de collecte de données, cela permet de déterminer avec précision quand les événements et les processus à analyser ont commencé, et quand ils se sont terminés.
- Le choix du projet comme unité d'analyse, permet de collecter des informations primaires et des informations secondaires d'une manière pratique. Nous avons remarqué que les personnes interrogées (éditeur de logiciel, client, acteur du réseau comme les sociétés de service) structurent et classifient l'information par projet, ce qui permet de retrouver facilement l'information.
- Le choix du projet comme unité d'analyse est pertinent pour analyser le phénomène de réseau, et en particulier la façon avec laquelle le projet est impacté par des stratégies managériales mais aussi par des événements extérieurs (Law et Callon, 1992). Cela permet d'identifier les acteurs impliqués dans le projet aussi bien que leurs caractéristiques, rôles et attitudes.

Notre recueil d'information doit nous permettre d'effectuer une comparaison de projet systématique et structurée, et de mettre en évidence les effets d'apprentissage et d'irréversibilité dont ils sont porteurs.

### **3.3 Une analyse processuelle :**

Nous recourons donc à une démarche méthodologique qui se justifie avant tout par notre question de recherche, que nous rappelons ici : « quels rôles jouent les premiers clients dans le processus de conception d'innovation logicielle radicale multi-acteurs, et comment les associer à ce processus ? ». Par conséquent, c'est moins l'image d'une situation à un moment donné que l'évolution de cette situation dans le temps qui nous intéresse. Une approche hypothético-déductive est bien adaptée pour comparer les changements intervenus sur un certain nombre de variables entre deux dates, notre objectif est de comprendre la dynamique de cette transformation. Au contraire, l'analyse d'un processus (Van de Ven et Poole, 1989) suppose de

suivre et d'analyser en temps réel les transformations afin d'élaborer une théorie fondée (Glaser & Strauss, 1967).

Notre travail s'inspire donc de l'approche processuelle de l'innovation au sens de Van de Ven (1989). Il s'inspire également des travaux d'Orlikowski (2005 ; Orlikowski et Baroudi, 1991) sur le développement des systèmes d'informations quant à la manière de structurer des données empiriques. Nous qualifions notre approche de « longitudinale » en suivant la définition proposée par Thiétart (Thiétart et al., 2003) : « les démarches longitudinales forment un ensemble d'analyses : focalisées sur l'étude de phénomènes au cours du temps. ». Toutefois, Forgues et Vandangeon-Derumez (2003) en donnent une définition adaptée de Ménard (1991), qui est beaucoup plus spécifique :

« 1) Les données recueillies portent au moins sur deux périodes distinctes ; 2) les sujets sont identiques ou au moins comparables d'une période à l'autre, 3) l'analyse consiste généralement à comparer les données entre (ou au cours de) deux périodes distinctes ou à retracer l'évolution observée ».

Or, dans le cadre de notre recherche, si nous nous attachons à observer le processus d'innovation au cours du temps, les sujets ne sont pas complètement « identiques ou au moins comparables d'une période à l'autre » (point 2).

Nous nous proposons d'étudier un processus d'innovation, au travers de deux projets réalisés pour les clients PressPro et Exinis. Le fait de n'étudier qu'une seule situation (innovation logicielle de « création automatisée de base de connaissance »), de ce genre se justifie par les raisons suivantes :

- la longue durée, nécessaire au déroulement de tout le processus d'implémentation d'une solution de « création automatisée de base de connaissance, utilisant la technologie du *text mining* (de 3 à 5 ans) » ne permet pas dans le cadre d'un travail de recherche en thèse, de suivre d'autres processus de même nature,
- la nouveauté de ce type de technologie entraîne la rareté des entreprises de logiciel susceptibles de réaliser des projets comparables,
- le caractère stratégique et très sensible de ce type de projet ne facilite pas l'analyse d'autres expériences similaires, réalisées par des entreprises œuvrant dans les

environnements très concurrentiels. Ces dernières ne souhaitent en général pas communiquer trop vite sur ces projets, ni surtout divulguer les inévitables difficultés rencontrées lors de leur mise en œuvre.

### **Biais – Subjectivité – Reconstitution a posteriori :**

Notre approche vise à « accéder à la logique des processus dans l'organisation, aux conditions du changement, aux causalités entre événements, aux changements d'objectifs et de contraintes » (Charue-Duboc, 2007). Une telle approche devant nous permettre d'éviter les biais de reconstitution a posteriori et décontextualisation des données (Journé, 2005) grâce à l'observation à intervalles de temps réguliers. Nous avons pour cela interrogé de manière régulière certains acteurs majeurs des projets étudiés, point sur lequel nous revenons au point 3.5 (Modalités de l'étude – Recueil de données). Le biais de rationalisation a posteriori peut aussi être évité grâce à la triangulation des sources « mais c'est plus complexe, quand il s'agit non pas d'oublis mais de jeux politiques qui poussent les acteurs à déformer volontairement les informations qu'ils livrent » (Journé, 2005). L'auteur poursuit en expliquant que « le biais de reconstruction a posteriori va de pair avec le problème de « décontextualisation » des données collectées ou construites par le chercheur (Dekker, 2003). Il risque alors de faire disparaître les contextes d'interprétation et d'action (Girin, 1990) dans lesquels les acteurs évoluaient au moment des faits. Toutefois, certains auteurs visent à restituer la richesse des contextes d'interprétation, en donnant la parole aux différents acteurs impliqués (Llory, 1996) » (Journé, 2005). Nous cherchons donc à éviter les biais évoqués, grâce à l'observation à intervalles réguliers des projets à l'origine de l'innovation et à la triangulation des sources, malgré les difficultés que cela comporte.

## **3.4 L'étude d'une situation emblématique de l'industrie du logiciel :**

Nous décrivons maintenant l'innovation dont nous avons étudié le processus de conception et de mise en œuvre, et justifions de son caractère emblématique.

### *3.4.1 L'étude de la conception d'un nouveau type de logiciel :*

En 2001, le centre de recherche XRCE avait la politique suivante : quand un projet de recherche aboutit à une innovation « tangible », une Business Unit est constituée en interne, chargée de développer et de commercialiser l'innovation. Ensuite, cette business unit devient autonome,

Xerox gardant les droits des logiciels (brevets) et revendant l'activité à une autre entreprise (généralement un éditeur de logiciel spécialiste de la Geide, gestion électronique de document). Des liens forts quoique partiellement informels sont conservés entre la business unit ainsi « essaimée », et le XRCE : il s'agit, entre autre, du maintien pendant un certain laps de temps de la présence de la business unit dans les locaux du XRCE, de la poursuite des relations interpersonnelles entre essaimés et chercheurs du XRCE, et, dans un premier temps, des accords avec XGS (Xerox Global Services), jouant le rôle d'intégrateur auprès des entreprises clientes (en juillet 2003, Temis a racheté l'activité « produits linguistiques » de Xerox).

Alors qu'au départ, nous pensions effectuer notre étude dans le cadre du centre de recherche XRCE de Xerox, notre travail a finalement été effectué, en grande partie, dans le cadre de l'entreprise Temis, qui a racheté et intégré l'activité « produits linguistiques » du groupe Xerox (i.e. l'équipe-projet XTS/Xelda et les droits d'exploitation des brevets afférents), en septembre 2003.

### 3.4.2 Les logiciels de « text mining », innovation radicale (ou innovation d'exploration) :

Temis dispose d'une gamme de logiciels dont les principaux sont IDE (extraction terminologique), IDK (catégorisation d'un document), IDC (*clustering*, regroupement de documents présentant des similitudes) et MServer (application fédérant les différents logiciels de Temis) qui a été remplacée en 2007, par Luxid. Pour chaque client, un développement personnalisé, qui complète IDE, doit être réalisé pour définir exactement les termes et concepts que celui-ci veut extraire : ce développement est appelé par Temis *cartouche de connaissance*. Chaque vente suppose également des développements personnalisés pour intégrer ces logiciels, eux-mêmes en évolution, au système d'information du client final.

Cette famille d'innovations peut être considérée comme une innovation radicale. D'une part elle s'appuie sur une combinaison d'algorithmes innovants d'analyse sémantique et d'analyse statistique. D'autre part, elle introduit une rupture dans les usages. En effet, l'offre que propose Temis à ses clients modifie les habitudes de travail des services qui l'adoptent, en permettant l'automatisation de l'analyse d'un texte : l'indexation et la catégorisation automatique (ou assistée par ordinateur), effectuées jusqu'alors manuellement, permettent ensuite un stockage électronique des documents (enrichis de métadonnées, comme la date de parution), et de la connaissance qu'ils contiennent (comme le thème principal d'un article), dans un système de gestion de base de

connaissance. Les utilisateurs peuvent ensuite effectuer des recherches parmi les documents et les connaissances ainsi stockées. Cette association entre un logiciel de *text mining*, et un système de gestion de base de connaissance constitue une innovation supplémentaire, qui aboutit à des solutions de « création automatisée de bases de connaissance » inédites.

### 3.4.3. L'accès à une situation emblématique : innovation modulaire multi-acteurs.

La situation étudiée est emblématique dans la mesure où elle illustre l'innovation née du couplage de logiciels eux-mêmes innovants, développés par deux start-up indépendantes. Nous avons ainsi pu étudier le lancement d'un nouveau type d'application logicielle (utilisant une technologie nouvelle, l'extraction terminologique exploitant des algorithmes sémantiques, combinée à celle encore peu répandue des bases de connaissance) dans un secteur où elle n'avait jamais été déployée (le domaine du *publishing*), et ce dans le cadre de start-up dédiées chacune à sa propre technologie innovante. C'est donc une illustration typique de l'innovation radicale dans l'industrie logicielle portée par des petites structures très spécialisées (Cusumano, 2004).

Le caractère emblématique de cette situation est renforcé par le fait que l'offre logicielle née de ce couplage, est susceptible de constituer les fondements d'une nouvelle plate-forme applicative dans le domaine du *text mining*, plate-forme sans équivalent à l'heure actuelle. Là encore, le caractère emblématique de la situation vient de ce que la logique de plate-forme est prégnante dans l'univers du logiciel (cf. chapitre 1 « L'industrie du logiciel »). L'intérêt de notre recherche est alors renforcé par le fait que la littérature, aussi bien sur le logiciel que sur la modularité, a très peu étudié le processus crucial d'émergence de telles plates-formes logicielles.

Nous avons donc eu l'opportunité d'analyser en temps réel la mise en œuvre d'un nouveau type d'application logicielle, dans le domaine du *text mining*, reposant sur le couplage entre deux logiciels eux-mêmes innovants, mais déjà mis en œuvre, séparément, auprès de quelques clients. La conception et la mise en œuvre de cette innovation ont été réalisées au travers de deux projets successifs, pour les clients PressPro puis Exinis France. Nous avons donc cherché à suivre la dynamique de ces projets, et l'émergence progressive d'une solution d'un type nouveau. Nous avons ainsi pu retracer l'évolution d'un processus d'implémentation d'une solution de *text mining* dans le domaine de l'édition.

### 3.5. Modalités de l'étude – recueil de données :

Nous relatons dans cette partie, les modalités de notre étude et le recueil de données que nous avons effectué.

#### 3.5.1 Modalités de l'étude :

Nous avons cherché à varier les sources de données (Yin, 2003) autant que possible pour assurer un maximum de fiabilité et de validité à ce que nous recueillions, sources qui nous ont permis une « triangulation de l'information », comme le préconisent Miles et Huberman (1984).

Selon Yin (2003), la collecte des données doit permettre de rassembler les preuves qui peuvent venir de six sources : documents, archives, entretiens, observation directe, observation participante, artefacts physiques. Nous présentons rapidement ci-dessous dans quelle mesure nous avons utilisé ces différentes sources, avant d'y revenir de manière plus précise par la suite :

**1. Documents :** nous n'avons pu accéder qu'à peu de documents écrits. Cela vient de ce que le type d'entreprise étudié de par sa nature, produit moins de documents écrits que, par exemple de grandes entreprises industrielles. Cela vient aussi du caractère confidentiel d'un grand nombre d'informations, dans un environnement très concurrentiel, sensible, comme nous l'avons déjà dit, aux questions de confidentialité. Toutefois, nous avons pu accéder, en octobre 2007, à un document particulièrement précieux pour notre travail de recherche, la thèse de doctorat en Informatique de F. Amardeilh (2007). Cette thèse porte sur le couplage des logiciels IDE/ITM de Temis et Mondeca, dont F. Amardeilh a eu la responsabilité au sein de Mondeca. Elle décrit précisément les logiciels qui ont été développés dans le cadre des projets PressPro et Exinis, ainsi que la méthodologie appelée « démarche OntoPop », qui vise à décrire l'ingénierie relative à la mise en œuvre du couplage entre les logiciels IDE de Temis et ITM de Mondeca. Enfin, nous avons pu accéder à certains documents comme le cahier des charges (en partie), et bien sûr des documents commerciaux, présentant l'offre et attestant de son évolution, ou indiquant la concrétisation de l'accord Temis/Mondeca par exemple.

**2. Archives:** peu dans notre cas, pour des raisons déjà explicitées qui tiennent à la jeunesse des entreprises étudiées et de leur offre.

**3. Entretiens** : ils ont clairement constitué la principale source de données de notre étude; il s'est agi d'entretiens récurrents avec des interlocuteurs clefs des deux projets étudiés, ainsi que ponctuels avec d'autres interlocuteurs. De par leur implication dans l'interaction avec nous, les interlocuteurs ont plus été « informants » que « répondants », au sens de Yin (2003), ce qui est critique dans la réussite d'une étude comme la nôtre. Nous insistons sur la régularité de ces entretiens, généralement bimestriels, dans le cas de l'account manager de Temis, notre principal interlocuteur.

**4. Observations directes et participantes** : au début de notre étude, nous avons participé à des réunions de projet internes chez Temis. Par la suite, nous avons assisté à des salons, et à des réunions commerciales associant Temis et Mondeca.

**5. Artefacts physiques** : les principaux artefacts physiques dans notre cas, sont constitués d'une part par les composants logiciels (génériques et personnalisés) développés par Temis, Mondeca et les autres acteurs des projets étudiés. Nous avons pu assister au fonctionnement des briques logicielles couplant IDE et ITM, appelée par Temis et Mondeca « plate-forme Ontopop », et que nous nommerons « passerelle OntoPop ».

Les informations recueillies proviennent de différentes sources : documents internes, observations réalisées lors de réunions, et entretiens semi-directifs principalement chez Temis et Mondeca. Néanmoins, nous n'avons pas eu les mêmes facilités d'accès aux documents internes (cahier des charges, spécifications) concernant l'innovation (certains documents relatifs aux projets, jugés confidentiels, sont restés en partie inaccessibles), ni à l'information détenue par les clients. Cela constitue une des limites de la méthodologie que nous avons employée, sur laquelle nous revenons au point 3.3.4. Par ailleurs nous avons cherché systématiquement à valider les informations recueillies, par des retours effectués aux acteurs du projet pour avis et validation. Parmi ces acteurs, ceux que nous avons interrogés régulièrement, notamment l'account manager et le OEM et partnerships manager de Temis, ont trouvé un intérêt immédiat à ces entretiens : celui de faire un point sur leur activité en général, et sur les projets en particulier, avec un acteur « neutre ». Cela a grandement facilité la validation croisée des informations recueillies auprès de ces personnes, et des autres personnes interrogées, ce qui constitue une triangulation supplémentaire.

### 3.5.2 Les données recueillies :

Les sources que nous avons utilisées, ont évolué avec le temps. Dans un premier temps (phases 1 et 2), nous nous sommes insérés dans les deux groupes de projet du XRCE, devenus Business Unit (regroupant de 20 à 30 personnes). Cette insertion nous a permis d'obtenir différents niveaux d'information sur la gestion du projet en interne, sur la relation fournisseur/client, sur la relation entre fournisseurs (éditeurs de logiciels, SSII) et sur le comportement du client face à l'innovation.

D'une manière plus précise, nous décrivons maintenant les différentes données recueillies au cours de notre étude :

- **En phases 1 et 2**, dans le cadre du XRCE de Xerox (cf. « construction de la place du chercheur » sur le terrain). Ces phases nous ont permis de nous familiariser avec les équipes, les technologies et les marchés de deux business units de Xerox, XTS/Xelda et AskOnce, les projets PressPro et Exinis n'ayant par ailleurs pas encore débuté. Le projet n'a donc pas constitué ici l'unité d'analyse, les phases 1 et 2 pouvant être considérées comme des phases préparatoires à la collecte de données elle-même :

- **Entretiens personnels** ouverts avec les dirigeants des business units de Xerox, XTS/Xelda d'une part, AskOnce d'autre part (business unit manager, marketing manager, responsable des relations OEM),
- Informations sur le fonctionnement de la business unit AskOnce, obtenues au travers de notre **participation en tant qu'observateur aux réunions** marketing, commerciales et de projet des unités,
- **Participation à des réunions de projet** XTS/Xelda avec un client de Xerox (EDF),
- **Participation à une présentation commerciale** présentant le projet développé par Temis pour un de ses tout premiers clients, Total, une application d'intelligence économique,
- **Entretiens avec un partenaire OEM** (Koltech, solution destinée aux DRH des entreprises).
- **Documents internes de Xerox**, relatifs aux projets.

- **En phase 3** (de fin 2003 à juillet 2007, à partir de l'essai des Business Units XTS/Xelda et AskOnce), les différentes sources de données, relatives aux projets PressPro et Exinis, auxquelles nous avons accédé furent les suivantes :

- **Entretiens personnels semi-directifs** avec des acteurs clés impliqués dans les projets : cela a constitué la principale source de données, comme nous l'avons dit. Chez Temis, nous avons entretenu des contacts récurrents avec certains d'entre eux (Vincent, ingénieur avant-vente/account manager pour le secteur du « *publishing* » ; Cyril OEM & Partnerships Manager ; Olivier, Core Product Technical Manager). Nous avons donc rencontré les principaux acteurs clés des projets étudiés chez Temis : nous avons interviewé, de manière régulière, environ tous les deux mois, l'account manager du secteur « *publishing* » de Temis, docteur en informatique ayant évolué vers des fonctions liées à l'avant-vente et à la responsabilité des comptes client dans le secteur de la presse et de l'édition, qui jouait un rôle de superviseur du projet, et gérait la relation avec les clients concernés, tout en apportant parfois une contribution technique. Régulièrement aussi, mais d'une manière plus espacée (tous les quatre mois à six mois), nous avons interviewé le responsable « OEM partnerships » de Temis ainsi que le responsable technique des « core products » de Temis, i.e. des différentes briques logicielles constituant l'offre de Temis.
- **Entretiens personnels semi-directifs**, et ponctuels, chez les **clients** :
  - un entretien avec la responsable de la documentation et chef de projet PressPro,
  - deux entretiens avec l'architecte innovation et membre de la DSI d'Exinis.
- Consultation du cahier des charges établis pour Exinis,
- Entretiens avec des managers de start-up logicielles dans le domaine de la recherche documentaire (P. Laval, directeur général de Sinequa ; Y. Mahé, Business Unit Manager de Documentum/AskOnce ; Stéphane Kolo, président et fondateur de Koltech);
- Participation à une réunion commerciale de présentation du partenariat OEM entre Thomson et Temis, par le CEO et le Directeur Commercial de Temis,
- Participation au salon professionnel Documation 2004,
- Documents commerciaux de Temis et de Mondeca,

- **En phase 4** (d'octobre 2007 à juin 2008), les différentes sources d'information, auxquelles nous avons accédé furent les suivantes :

- **Thèse de doctorat** en Informatique réalisée par la responsable de couplage (Florence Amardeilh, 2007).
- **Entretiens personnels semi-directifs** récurrents avec des acteurs clefs impliqués dans le projet chez **Temis** : avec l'ingénieur avant-vente et account manager; tous les deux mois environ ; et deux entretiens avec l'OEM & Partnerships Manager.
- **Entretiens personnels semi-directifs** avec des acteurs clefs impliqués dans le projet chez **Mondeca** : trois entretiens avec la responsable du couplage Temis/Mondeca, trois entretiens avec le PDG de Mondeca, Jean Delahousse, un entretien avec le chef de projet PressPro et Exinis ;
- **Deux entretiens personnels** semi-directifs avec l'architecte innovation, membre de la DSI d'**Exinis**.
- **Participation à la présentation commerciale** de la suite logicielle Luxid de Temis en présence des principaux clients et partenaires éditeurs de logiciel de Temis, en juin 2008.

Notre collecte d'information a été centrée sur les projets PressPro et Exinis, et la relation fournisseur/client dans le cadre de ces projets, mais comme le montrent les travaux du CSI, et Christensen et Rosenbloom (1995), les réseaux jouent un rôle primordial dans la réussite commerciale d'une innovation. L'entreprise qui innove doit savoir créer des alliances (accords entre éditeurs de logiciel, ou éditeurs de logiciel/SSII) qui lui permettront de créer l'environnement nécessaire au succès de son innovation (Lenfle, 2001). Nous avons donc aussi tenté, lors de notre collecte de matériau, de rechercher les informations relatives à la façon dont l'éditeur de logiciel innovant s'insère dans un réseau, en choisissant les bonnes alliances, grâce notamment à nos échanges avec le « OEM & Partnerships manager » de Temis.

#### **Bilan du nombre d'entretiens par projet :**

**Projet PressPro** (09/2003 à 11/2006) : 29 entretiens.

→ 24 entretiens avec des personnes de Temis :

- 17 entretiens avec l'ingénieur avant-vente/account manager
- 5 entretiens avec le « OEM & partnerships manager »,
- 2 entretiens avec le « Core product technical manager ».

→ 4 entretiens, a posteriori, avec des personnes de Mondeca :

- 1 entretien avec le président,
- 2 entretiens avec la responsable de couplage,

- 1 entretien avec le chef de projet.

→ 1 entretien avec la chef de projet de PressPro.

**Projet Exinis** (06/2004 à 06/2008) : 37 entretiens.

→ 27 entretiens avec des personnes de Temis :

- 20 entretiens avec l'ingénieur avant-vente/account manager,

- 5 entretiens avec le « OEM & partnerships manager »,

- 2 entretiens avec le « Core product technical manager ».

→ 6 entretiens avec des personnes de Mondeca :

- 2 entretiens avec le président,

- 2 entretiens avec la responsable de couplage,

- 2 entretiens avec le chef de projet.

→ 4 entretiens avec l'architecte innovation d'Exinis,

### **Le guide d'entretien :**

Pour structurer notre processus de collecte de données, dans la phase 3, nous avons utilisé une liste d'items comme *guideline*, ou lignes directrices, pour nos guides d'entretien semi-directifs (Möller, 1986). Cette liste d'items était la suivante:

- les **tâches** à réaliser, leur déroulement et leur enchaînement,
- les **coûts** liés à ces tâches,
- les **interactions** entre éditeurs de logiciel, et entre éditeurs de logiciel et client (personnes impliquées avec leurs caractéristiques, objectifs, durée, intensité, atmosphère, résultats de l'interaction),
- le **leadership** du projet,
- les événements inattendus et difficultés du projet,
- les **inflexions** du projet,
- l'**apprentissage** et la **capitalisation** permis par le projet, la définition d'une ingénierie propre à l'intégration d'une telle innovation,
- les effets **d'irréversibilité** relatifs à l'innovation elle-même, notamment les composants logiciels développés, qu'ils soient génériques ou personnalisés.

Notre objectif était d'obtenir un flux « naturel » d'information sur chaque projet, plus que de suivre un flot de questions rigide et détaillé. Cette posture plaçait les interviewés dans une position d'« informant », opposée à celle d'« interrogé », répondant à des questions prédéfinies (Yin, 2003) et préservait le flot chronologique (Miles & Huberman, 1994). Notre rôle pouvait être comparé à celui d'un enquêteur, un coordinateur, un modérateur adoptant un ton neutre pendant l'entretien (Akrich, Callon et Latour, 1988).

### **Les aspects pratiques du processus d'entretien :**

Pratiquement tous les entretiens se sont déroulés en face à face (quelques uns ont eu lieu par téléphone) dans les locaux de l'entreprise de la personne interviewée. Ils ont été enregistrés (avec l'autorisation de l'interlocuteur) et retranscrits. Nous avons aussi pris des notes pendant chaque entretien. D'abord, pour éviter de perdre le matériau en cas de problème technique d'enregistrement, ensuite pour garder une trace des données quand le répondant demandait d'arrêter l'enregistrement pour des aspects « sensibles » (trois cas), et écrire des commentaires personnels pendant l'entretien. La durée de chaque entretien a varié entre 45 minutes et 2 heures, selon l'évolution des projets et selon la nature et le niveau d'implication (niveau de connaissance du projet) de l'interviewé dans ces projets. La durée moyenne des entretiens étant de 1h15, le temps cumulé consacré à ces **66 entretiens** est évalué à **82 heures**.

### *3.5.3 Les limites de la mise en œuvre de la méthodologie retenue :*

Nous focalisant sur un processus de coopération et de partenariat entre deux entreprises, il était essentiel de collecter des informations de chaque côté de la relation. Yin (2003, p. 21) « si vous réalisez votre étude en étudiant seulement une organisation, vous ne pouvez pas tirer des conclusions précises, sur des partenariats interorganisationnels. C'est un « vice » dans votre research design. » Or comme, nous l'avons vu plus haut, le recueil d'information auprès des clients, tout comme celui réalisé auprès de Mondeca, a été plus limité et plus tardif que celui auprès de Temis. Cela constitue donc un biais dans notre analyse processuelle, car nous n'avons pas toujours pu, de ce fait, confronter les différentes visions des acteurs des projets en temps réel. Nous exposons ci-après les raisons de ces limites et décalages temporels.

L'accès au terrain est un processus progressif, qui nous a permis d'acquérir la confiance de nos différents interlocuteurs, et des organisations auxquelles ils appartenaient. Ainsi le travail réalisé

en phase 1, nous a permis d'accéder au terrain de recherche en phase 2, et surtout en phase 3. Par la suite, l'accès aux acteurs-clefs chez Mondeca, a été rendu possible par le travail réalisé auprès des acteurs concernés chez Temis, mais aussi chez Exinis. Nous insistons à nouveau sur le fait, que les start-up logicielles que nous étudions, sont soumises à de fortes turbulences technologiques et de marché, ainsi qu'à une concurrence à la fois intense et difficile à cerner, car provenant d'entités nouvelles, souvent de petites tailles, disséminées dans le monde entier. C'est pourquoi, nos interlocuteurs chez Temis étaient logiquement réticents au départ à ce que nous interrogeons leurs clients, voire même leur partenaire Mondeca. Les technologies de Temis et de Mondeca sont différentes, mais il peut y avoir des recoupements : ainsi la suite logicielle Luxid de Temis, lancée en juin 2008, offre-t-elle la possibilité de réaliser un système de gestion de base de connaissance, certes bien moins évolué que celui proposé par Mondeca. En ce qui concerne les clients, Temis ne souhaitait pas au départ que nous interagissions avec eux, les incertitudes étant fortes autour de projets nouveaux comme ceux que nous avons étudiés, et les risques d'échec réels. C'est pourquoi, Temis voulait « contrôler » au maximum notre relation avec ses clients. Ensuite, nous avons pu les interroger, ponctuellement, mais cela a induit un décalage temporel entre la vision que nous avons recueillie auprès de Temis et celle que nous avons recueillie auprès de ses clients. Le problème est le même, en ce qui concerne Mondeca. En effet, il existe généralement des enjeux concurrentiels entre éditeurs de logiciel dans un même champ d'application, et donc des préoccupations relatives à la confidentialité chez chacun d'entre eux. C'est une des raisons pour lesquelles, nous n'avons pu accéder aux acteurs des projets chez Mondeca, que tardivement au cours de notre étude. En conclusion, il peut apparaître que notre position sur le terrain n'a pas été idéale pour collecter les données nécessaires à notre recherche, du fait de l'accès tardif à certains acteurs des projets étudiés. Nous en avons conscience, mais nous considérons que cela est dû à la nature de l'industrie où s'est déroulée notre recherche, et qu'en tout état de cause, cette place de chercheur devait être construite progressivement, et qu'il eût été vain d'attendre de pouvoir accéder à tous les acteurs concernés pour débiter réellement notre recherche. Nous estimons en effet que c'est bien parce que nous avons débuté notre collecte de données auprès de Temis, que nous avons pu ensuite obtenir la confiance des acteurs des projets chez Temis et que nous avons pu la poursuivre auprès de leurs clients et de Mondeca. Une autre raison à cela, est bien entendu l'excellente entente qui s'est développée au fil du temps et des collaborations entre les acteurs des projets de Temis, Mondeca et Exinis. En conclusion, si l'analyse processuelle nous paraît effectivement être la démarche la mieux adaptée à notre recherche, nous estimons que les limites rencontrées liées à sa mise en œuvre découlent

logiquement de la nécessité d'une construction progressive de la place du chercheur sur le terrain, dans le cas d'une innovation logicielle radicale multi-acteurs.

#### 3.5.4 Conclusion de la partie « Modalités de l'étude - Recueil de données » :

Sur la base des données que nous avons collectées, nous avons rédigé deux monographies, l'une relative au projet PressPro, l'autre relative au projet Exinis. Celles-ci permettent de suivre la progression de ces deux projets, et en conséquence la trajectoire de l'innovation étudiée, de l'expression de l'idée d'une architecture nouvelle, à sa mise en œuvre réussie chez Exinis.

Nous avons décrit cette collecte de données en nous efforçant de suivre les préceptes de Popper (1973) : « ce que l'on peut appeler objectivité scientifique repose uniquement sur la tradition critique qui, en dépit des résistances rend souvent possible la critique d'un dogme qui prévaut (Girin, 1979). Autrement dit, la subjectivité soumise au contrôle collectif peut permettre [...] de juger plus ou moins plausible telle reconstruction des logiques des comportements des acteurs et contribuer à évaluer le pouvoir explicatif des constructions théoriques et à sélectionner celles qui sont meilleures que les autres ». On comprend alors l'importance des instances de contrôle mais aussi de la diffusion des résultats de la recherche à la communauté scientifique. Le chercheur en gestion doit exposer ses résultats mais également rendre lisible sa trajectoire de recherche. Les interactions avec le terrain conduisent à faire évoluer cette recherche dans le temps. C'est en explicitant les conditions de sa production que l'énoncé devient falsifiable : ainsi pour Yin (1994) « le principe est de permettre à un observateur externe –le lecteur de l'étude de cas, par exemple, de suivre le cheminement de n'importe quelle preuve présentée, des questions de recherche initiales aux conclusions ultimes du cas. De plus, cet observateur externe doit être capable de retracer les étapes dans n'importe quelle direction (des conclusions vers les questions de recherche initiales, ou des questions vers les conclusions) ». On rejoint là la position de G. Dumézil pour qui « la méthode, c'est le chemin après qu'on l'a parcouru » (cité dans Hatchuel, 1994, p. 71).

## 4. CONCLUSION : UNE OPPOSITION OBSERVATION/INTERVENTION A RELATIVISER.

Nous voulons revenir sur le concept de « recherche clinique » qui « peut se définir comme l'interaction instituée entre le chercheur et son terrain d'étude » (Girin, 1981 ; Berry M., 1981).

Elle consiste à rentrer dans les organisations pour les aider à formuler et à résoudre les problèmes qu'elles ont ressentis ou formulés. La typologie établie par Schein (1987) affirme que le chercheur-clinicien se situe dans le cadre du modèle de consultation dynamique dans lequel il travaille avec l'entreprise à la formulation du problème et à la mise en œuvre de la solution. L'appellation « clinique » renvoie ainsi à ceux qui « aident des professionnels, ceux qui sont impliqués auprès des personnes » (Schein, op. cité). La recherche clinique se fonde sur un problème, plus ou moins explicité par l'organisation. Lenfle (2001) note que la recherche peut s'engager sur la base d'interrogations de l'organisation sur des points qui ne posent pas nécessairement, ou pas encore, problème. La recherche est alors une situation de co-production de connaissances sur une question. Un processus d'interaction va s'engager qui va permettre d'explorer à la fois la nature du problème (la question posée par l'entreprise), et la solution la mieux adaptée (la théorie du chercheur). Dans le cas d'une recherche interactive, « la nature de l'étude évolue dans le temps » (Lundin & Wirdenius, 1990) dans la mesure où l'objectif « est de comprendre le système étudié et de laisser cette compréhension évoluer dans l'interaction entre le chercheur et les acteurs du système ». « L'opportunisme méthodique » promu par Girin (1981), est alors une ressource normale d'une recherche interactive.

Charue-Duboc (2007) observe que depuis vingt ans en France, des chercheurs en sciences de gestion se sont engagés dans des démarches d'intervention, qui ont fait l'objet de réflexions méthodologiques et épistémologiques (Moison 1984, Hatchuel & Molet 1986, Girin 1983, 1990, Martinet 1990, Chanal, Lesca & Martinet 1997, Koenig 1993, David 2000). L'auteur (Charue-Duboc, 2007) explique qu'à l'intérieur de ces familles des différences entre les pratiques existent et des termes variés sont utilisés : recherche clinique, recherche ingénierique, recherche-intervention [...]. Un point commun essentiel à ces approches est de considérer l'interaction entre le chercheur et son terrain comme un moyen sur lequel le chercheur doit s'appuyer dans son analyse et non comme un biais ou une faiblesse méthodologique. [...] Un second trait commun est que le point de départ d'une telle recherche est un problème ressenti par les acteurs du terrain. En cela, cette tradition se démarque de la recherche action telle que préconisée par Lewin (1951) qui considérait que l'objectif d'analyse et de compréhension porté par le chercheur était prioritaire par rapport aux objectifs opérationnels des praticiens. Les chercheurs du Centre de Recherche en Gestion (CRG) au sein duquel nous avons effectué notre recherche, considèrent que le fait que le praticien ait identifié un problème sur lequel il souhaite solliciter un chercheur est un point d'appui essentiel. La démarche clinique consiste alors à construire une relation avec une organisation sur la base d'une demande formulée en son sein. L'intérêt de partir d'une

demande est d'abord que la question posée porte sur des enjeux réels ressentis par les acteurs. De plus, les praticiens sont plus enclins à s'engager dans une interaction avec un chercheur, voire à lui laisser prendre une place d'intervention, s'ils sont confrontés à un problème. L'étude de telles évolutions suppose alors d'inscrire les démarches sur la durée. Nous considérons alors que la démarche que nous avons suivie pour effectuer notre travail de recherche est une démarche clinique, en ce qu'elle considère l'interaction entre le chercheur et son terrain comme un moyen sur lequel le chercheur doit s'appuyer dans son analyse, et en ce que son point de départ est un problème ressenti par les acteurs du terrain.

Toutefois, il se pose, dans notre cas, la question de la distinction entre étude clinique et observation (participante ou non participante). En simplifiant, nous pouvons dire qu'en phase 1, nous étions acteurs, en phase 2 observateur participant, et en phases 3 et 4, qui constituent le cœur de notre travail de recherche, observateur passif. En phase 2, nous avons pratiquement obtenu, pendant quelques mois, la position qualifiée d' « idéale » par Girin (1990), celle d' « étranger intégré » (dans la business unit AskOnce). Mais comme l'observe Raulet-Croset (2003), la position du chercheur n'a rien de définitif, comme nous l'avons compris après la vente de ses business units par Xerox, et la remise en cause de notre travail de recherche qui s'en est suivie.

Par la suite, en phase 3 et 4, le caractère explicitement actif de notre rôle dans la vie de l'entreprise Temis s'est effacé, pour laisser la place à une observation apparemment passive. Toutefois, nous considérons que cette opposition entre observation et participation doit être relativisée (Hatchuel 1997). Tout d'abord parce que l'introduction d'un chercheur dans une entreprise constitue en soi une perturbation. Ainsi, « à simplement vouloir observer, on agit sur la réalité que l'on voudrait saisir, et cette réalité en retour agit sur la dynamique de la recherche » (Girin 1990). Les analyses et conclusions du chercheur ne sont pas non plus neutres. Nous pensons alors que le chercheur doit faire partie du milieu qu'il étudie, considérant que son rôle doit être de réfléchir à sa position et de rendre compte de son « intervention ». Ainsi, Les travaux de J. Favret-Saada (1977) montrent que, quand bien même le chercheur se présente comme neutre, ses interlocuteurs lui affectent une place. Elle souligne également que certains phénomènes ne peuvent être analysés sans prendre un rôle dans une situation. Cette auteure considère alors l'interaction comme centrale dans la construction d'une compréhension de la situation. Raulet-Croset (2003) observe qu'il est « rare de pouvoir avoir accès à l'organisation sur une longue durée sans faire un apport, sous quelque forme que ce soit, à l'organisation (sans forcément d'ailleurs répondre à une demande explicite) ». C'est ce que nous avons pu constater

dans le cas de notre étude, puisque les personnes interrogées sur les projets en cours (phases 3 et 4), trouvaient pour certaines d'entre elles en tout cas, un intérêt à leurs entretiens réguliers avec nous : la possibilité de faire le point avec un acteur neutre, extérieur à leur entreprise, leur permettant de prendre du recul par rapport à leur activité. Ainsi, reprenant à notre compte l'opportunisme méthodique prôné par Girin (« La relation doit également être gérée dans le temps. J. Girin (1989) parle d'opportunisme méthodique en insistant sur le fait que les évolutions du terrain peuvent être considérées de façon positive comme une opportunité à saisir plutôt que comme une entrave à la recherche. », Raulet-Croset, 2003) avons-nous en permanence essayé d'évaluer notre position dans les entreprises, et, de comprendre la façon avec laquelle nous étions perçu et l'intérêt que notre travail pouvait susciter chez nos interlocuteurs. Cette attention nous a paru nécessaire à la fois pour préserver notre accès au terrain, aussi bien que pour analyser les discours des personnes interrogées à propos des projets étudiés.

Notre travail correspond donc, de notre point de vue, à une étude clinique, avec interaction directe et organisée avec un certain nombre d'acteurs clefs chez les fournisseurs et les clients. Cette étude clinique est fondée sur l'analyse processuelle d'une situation emblématique. Notre projet s'appuie donc sur une approche inductive. A la différence d'une approche déductive qui se nourrit de théories pour valider un point de vue ou bien une hypothèse théorique du départ, la méthode inductive démarre par l'analyse des phénomènes de gestion et l'observation de la réalité. Dans une deuxième phase, la méthode s'appuie sur les théories existantes ou d'autres éléments analytiques pour essayer de décrire les phénomènes observés. L'application des théories aux résultats de l'observation nous amène aux conclusions de la recherche : il s'agit d'une part de valider les hypothèses suggérées par les théories et d'autre part, de valider l'utilisation des théories pour l'explication des phénomènes. Cette approche offre des parcours de recherches complémentaires et des scénarios futurs. Enfin, la démarche inductive conduit à généraliser une série d'observations empiriques à travers une loi ou un modèle. On observe, on localise, on dénombre, on mesure, on classe. Puis, on recherche des régularités qui permettent de faire progresser la connaissance ; la construction des modèles permet d'explicitier ces régularités. En effet, si notre démarche est fondée sur l'approche inductive, certaines des conclusions ou conjectures dégagées se rapportent à un compromis entre induction et abduction. Par abduction, nous entendons : l'opération qui n'appartenant pas à la logique, permet d'échapper à la perception chaotique que l'on a du monde réel, par un essai de conjecture sur les relations qu'entretiennent effectivement les choses. Alors que l'induction vise à dégager des régularités

indiscutables, l'abduction consiste à tirer de l'observation des conjectures qu'il convient ensuite de tester et de discuter (Koenig, 1993).

**Chapitre 4 :  
Emergence d'une innovation au  
travers de deux projets**

## SOMMAIRE DU CHAPITRE 4 :

<b>Chapitre 4 : Emergence d'une innovation au travers de deux projets</b>	<b>135</b>
<i>Sommaire du Chapitre 4 :</i>	136
1. <i>Les éditeurs de logiciel Temis et Mondeca :</i>	139
1.1 Temis :	139
1.2 Mondeca :	141
2. <i>Les outils de l'annotation sémantique et de l'extraction terminologique :</i>	142
2.1 Le web sémantique :	142
2.2 Constats issus d'un pré-projet en 2002 :	143
2.3 Un outil d'extraction d'information : <i>Insight Discoverer Extractor™</i> de Temis.	146
2.4 Synthèse et ouverture sur l'innovation à développer :	151
3. <i>Le projet PressPro :</i>	153
3.1 L'origine du projet :	153
3.2 Description et chronologie du projet	154
3.2.1 La solution à mettre en place :	154
3.2.2 La chronologie du projet :	155
3.3 Les phases amont jusqu'à la signature du contrat définitif :	156
3.3.1 Première étape avant la signature d'un premier contrat :	156
3.3.2 Un premier contrat en octobre 2003 : la rédaction des spécifications du projet.	158
3.3.3 Le contrat définitif entre les trois éditeurs signés en mai 2004 :	162
3.4 L'organisation du projet à partir de mai 2004 :	164
3.4.1 Les acteurs du projet :	164
3.4.1.1 Chez Temis :	164
3.4.1.2 Chez Mondeca :	165
3.4.1.3 Chez Xyleme :	166
3.4.1.4. Chez PressPro :	166
3.4.2 Un tissu d'interactions dense :	167
3.4.3 Les étapes du projet :	170
3.4.3.1 Les briques logicielles de Temis :	170
3.4.3.2 L'application documentaire par PressPro, de mai à décembre 2004 :	172
3.4.3.3 Les briques logicielles de Mondeca :	173
3.4.3.4. Le système intégré (à partir de décembre 2004) :	174
3.4.3.4. Le système intégré (à partir de décembre 2004) :	174
3.5 Synthèse –Résumé du projet PressPro :	179
4. <i>La passerelle logicielle OntoPop et la notion de RAC (règles d'acquisition des connaissances) :</i>	185
4.1 Introduction :	186
4.2 Une passerelle pour l'annotation sémantique et le peuplement d'ontologie :	187
4.3 La formalisation des Règles d'Acquisition de Connaissance	189
4.3.1 L'importance du contexte dans les arbres conceptuels :	189
4.3.2 Méthode et règles d'exploration contextuelle fondement des RAC :	191
4.3.3 Des règles d'exploration contextuelle aux règles d'acquisition de connaissance :	192
4.3.4 Définition des constituants d'une Règle d'Acquisition de Connaissance : PressPro	192
4.3.4.1 Introduction :	192
4.3.4.2 Les composants d'une RAC :	193

4.3.5	Edition des Règles d'Acquisition de Connaissance grâce à un éditeur de règle :	195
4.3.6	Conclusion :	195
5.	<i>Fonctionnement de la passerelle OntoPop :</i>	196
5.1	Cycle vie des RTO dans la démarche OntoPop :	196
5.1.1	L'analyse linguistique :	196
5.1.2	L'application des Règles d'Acquisition de Connaissance :	196
5.1.3	L'enrichissement des lexiques linguistiques de l'outil d'extraction de l'information via le module « mise à jour des lexiques » :	198
5.2	Fonctionnement des modules logiciels d'OntoPop :	198
5.2.1	Le module d'annotation sémantique et de peuplement ontologique :	198
5.2.1.1	La transformation :	199
5.2.1.2	La consolidation :	199
5.2.1.3	La validation :	202
5.2.2	La maintenance des lexiques et autres ressources linguistiques :	202
5.3	Les composants de la passerelle logicielle OntoPop :	203
6.	<i>La méthodologie OntoPop.</i>	204
6.1.	Introduction :	204
6.2	Présentation générale de la méthodologie OntoPop :	205
6.3	Les différentes phases de la méthodologie OntoPop :	208
6.3.1	La Phase d'Etude :	208
6.3.2.	La Phase de Structuration :	213
6.3.2.1	Modélisation de l'ontologie du domaine :	213
6.3.2.2	Construction des cartouches linguistiques :	214
6.3.3	La Phase de Couplage :	218
6.3.4	La Phase de Validation :	220
6.3.5	La Phase de Mise en Service :	222
7	<i>Le projet Exinis :</i>	224
7.1	L'origine du projet Exinis :	224
7.2	Description et chronologie du projet :	225
7.3	Les phases amont jusqu'à la signature du contrat définitif :	228
7.3.1	Premiers contacts en avril 2004 :	228
7.3.2	Un premier contrat pour un démonstrateur en juin 2004 :	228
7.3.3	Présentation du démonstrateur à Exinis US :	229
7.3.4	Le contrat définitif signé en mai 2005 :	230
7.4	Les acteurs et l'organisation du projet (à partir de mai 2005) :	231
7.4.1	Les acteurs du projet chez Exinis :	231
7.4.2	Les acteurs du projet chez Temis :	233
7.4.3	Les acteurs du projet chez Mondeca :	234
7.4.4	Les acteurs du projet chez 4D Concept :	234
7.4.5	L'acteur du projet chez Unilog :	234
7.4.6	Les réunions de projet organisées par l'architecte innovation d'Exinis :	235
7.5.1	Développement de la version V0 de la solution (mai 2005-février 2006):	235
7.5.2	Relance du projet et finalisation de la version V1 (juin 2006 - juin 2008) :	239
7.6	Synthèse - résumé : Principales différences entre les projets PressPro et Exinis.	241
8.	<i>Conclusions et apports des projets PressPro et Exinis :</i>	242

Les éléments méthodologiques étant posés, nous présentons les monographies des projets PressPro et Exinis (noms d'entreprises que nous avons modifiés par soucis de confidentialité). Nous avons effectué le suivi de ces deux projets sur plusieurs années afin de réaliser l'analyse processuelle que nous avons décidé de conduire. Nous avons choisi de relater d'une manière relativement brute le fil de ces projets, tel que nous l'avons constitué progressivement, afin de fournir au lecteur les éléments lui permettant de reconstituer par lui-même le déroulement de ces projets, et in fine le processus d'innovation lui-même.

Nous avons donc étudié le cas d'une innovation logicielle, un système de gestion et d'enrichissement automatisés de base de connaissance, que nous appelons par souci de simplification « système de création automatisée de base de connaissance », destinée aux entreprises. Cette innovation s'inscrit dans le cadre du développement de ce que les spécialistes nomment le « web sémantique », et vise à résoudre le problème de la gestion (indexation, classement, recherche, mise à disposition d'utilisateurs divers) d'un très grand nombre de documents textuels (articles de presse, articles scientifiques, textes de lois, brevets, dépêches d'agence,...). Les projets étudiés combinent pour la première fois différentes briques logicielles indépendantes permettant l'extraction d'information (moteur d'extraction IDE de l'éditeur de logiciel Temis), et la gestion de la connaissance (système de représentation de la connaissance ITM, de l'éditeur de logiciel Mondeca). Dans un premier temps, c'est une solution logicielle innovante qui se structure principalement à l'occasion du projet conduit chez PressPro, et du développement de la « plate-forme logicielle OntoPop » (terminologie employée par Mondeca et Temis), puis dans un deuxième temps, c'est une méthodologie visant à décrire la mise en œuvre d'une telle « plate-forme » qui est développée. Nous observons ainsi le passage d'une situation dans laquelle différents éditeurs de logiciel s'appuyant sur des travaux de recherche novateurs, proposent de nouveaux concepts de logiciel non couplés, à une situation dans laquelle une solution logicielle, dont la structure et la mise en œuvre sont clairement définies, peut être commercialisée de manière récurrente. Ainsi, cette offre sera, par la suite, déployée chez d'autres clients, comme l'AFP, et Press Index.

Nous présentons donc dans un premier temps les deux start-up qui portent cette innovation, Temis et Mondeca, puis nous décrivons les développements technologiques réalisés dans le cadre du web sémantique. Par la suite, nous retraçons le déroulement, sur plusieurs années, des ces deux projets, en insistant sur le rôle des acteurs, et leurs interactions (éditeurs de logiciel, clients, sociétés de services informatiques en charge de l'intégration de la solution), lors des différentes phases de chaque projet, ainsi que sur l'émergence progressive de l'innovation, et de la méthodologie nécessaire à sa mise en œuvre. Les deux projets ont permis à Temis et Mondeca de définir une innovation logicielle destinée au secteur de la presse et de l'édition. Grâce à l'apprentissage réalisé sur le projet PressPro, l'architecture de la plate-forme logicielle OntoPop a été définie, tout comme les RAC permettant le couplage entre logiciels d'extraction d'information et de représentation des connaissances, couplage qui constitue, avec la capacité à définir une cartouche de connaissance permettant d'extraire les informations pertinentes d'un document écrit, un des éléments critiques de la mise en œuvre de la « plate-forme OntoPop ». A ce propos, nous préférons retenir le terme de passerelle logicielle OntoPop, qui correspond à notre sens mieux à la nature de ce composant logiciel. Ce projet PressPro a aussi permis, avec le travail effectué pour concevoir la version V0 de la solution destinée à Exinis, la définition d'une méthodologie, méthodologie qui a ensuite été appliquée aux dernières phases du projet Exinis.

C'est pourquoi nous intercalons la présentation de la passerelle logicielle et de la méthodologie OntoPop, entre les monographies des projets PressPro et Exinis.

## **1. LES EDITEURS DE LOGICIEL TEMIS ET MONDECA :**

### **1.1 Temis :**

Temis a été fondée en septembre 2000, par une équipe de dirigeants, chercheurs et consultants d'IBM, pour développer et commercialiser des solutions innovantes de *text mining*. Ainsi, Temis propose des outils pour l'organisation de collections de documents et pour l'extraction d'information. Ces outils doivent permettre aux entreprises d'optimiser le traitement de leurs flux d'information, en transformant du texte libre en données analysables. Le *text mining* est, en effet, un ensemble de traitements informatiques consistant à extraire des connaissances selon un critère de nouveauté ou de similarité dans des textes. Dans la pratique, cela revient à mettre en algorithmes un modèle simplifié des théories linguistiques dans des systèmes informatiques d'apprentissage et de statistiques. Il s'agit donc d'une innovation de par ses fondements

technologiques nouveaux, et même d'une innovation radicale, car la technologie du *text mining* perturbe les habitudes de travail des chercheurs, spécialistes de l'intelligence économique, documentalistes et autres utilisateurs, en automatisant l'organisation de collections de documents et l'extraction des informations dont ils ont besoin.

Temis a signé, en 2000, un accord initial de licence avec Xerox pour intégrer Xelda, son moteur d'analyse linguistique, issu du Centre de Recherche Européen XRCE de Xerox, basé à Grenoble. Cet accord a ensuite permis, en 2002, à Temis d'acquérir l'activité «produits linguistiques» du groupe Xerox, et d'intégrer l'équipe-projet de Xerox (douze personnes) en charge des logiciels Xelda et XTS (logiciel d'extraction terminologique). Grâce à cette acquisition, Temis a renforcé son équipe de développement avec un groupe d'experts en linguistique et a complété sa gamme avec les produits linguistiques de Xerox, Xelda et XTS. La société employait, en décembre 2007, plus de 50 personnes et possédait des filiales en France, Allemagne, Italie, Royaume-Uni et aux Etats-Unis. Temis se présente comme leader sur la technologie du *text mining* (appelé aussi parfois «*text analytics*»), et comme la première société à avoir spécialisé ses outils à la fois sur des problématiques spécifiques aux entreprises (l'Intelligence Economique, l'Intelligence Scientifique, l'analyse de la relation client et la Gestion des Ressources Humaines) et sur des secteurs précis (sciences de la vie, médias, automobile). C'est un acteur internationalement reconnu, dans un secteur qui en compte peu, certains ayant été rachetés récemment par de grands éditeurs de logiciel comme Fast par Microsoft en 2008, et InXight par Business Objects en 2007. Cet intérêt témoigne des espoirs placés dans ce que les spécialistes appellent le « marché de l'accès à l'information en entreprise» ou «*enterprise search*», et que S. Ballmer, CEO de Microsoft, a chiffré à 13 milliards de dollars en 2006 par an (Ballmer, 2006).

Les premières ventes de Temis ont été réalisées dans des secteurs variés (tri automatisé de CV pour la DRH d'une banque, intelligence économique pour une compagnie pétrolière, veille technologique pour un laboratoire pharmaceutique). Les dirigeants de Temis étaient persuadés du potentiel commercial de leur offre dans le domaine de la presse et de l'édition (le «*publishing*»), et souhaitaient investir ce secteur. Secteur dans lequel Mondeca avait déjà commencé à travailler. L'opportunité leur en a été donnée en 2003, lorsqu'ils ont été contactés, ainsi que Mondeca, par la responsable de la documentation du groupe PressPro. Ce contact faisait suite à une première ébauche de projet, réalisé pour un laboratoire pharmaceutique en 2002, qui avait demandé à Temis et à Mondeca de réaliser une solution de couplage entre l'outil d'extraction d'information IDE, et le système de gestion de base de connaissance ITM. Le travail réalisé par ces deux

éditeurs de logiciel n'avait pas abouti : il les a tout de même conduits à passer un accord de partenariat OEM pour de futures réalisations, et leur a permis d'entrevoir un certain nombre de difficultés inhérentes à une telle innovation. Le projet conduit pour PressPro (projet de « création automatisée de base de connaissance ») a permis ensuite, en 2004, à Temis et Mondeca d'aborder un projet de plus grande envergure pour le groupe Exinis, éditeur international (« fournisseur d'information professionnelle »).

## 1.2 Mondeca :

La société Mondeca a aussi été créée en 2000, par Jean Delahousse, avec l'idée de faire de « référentiels de connaissance », au moment où les normes sur le « Web Sémantique » (cf. point suivant) ont commencé à apparaître. Jean Delahousse a décidé de créer Mondeca après une expérience commerciale et managériale d'une dizaine d'années dans l'édition de logiciel financiers. Mondeca s'est spécialisé dans la gestion d'ontologies, de thésaurus, de bases de connaissance, et de manière générale dans l'organisation des sujets et l'indexation des contenus (cf. Annexe, page 359). Dès 2002, Mondeca a intégré, par l'intermédiaire d'un chercheur qui jouait le rôle d'expert ontologue (« consultant »), le groupe de travail OWL (OWL est un langage informatique permettant de définir des ontologies). Cet expert a travaillé sur les normes, participant à différentes activités de standardisation dans le cadre de la *W3C Semantic Web Activity : WebOnt (OWL) Working Group* (2002-2003), *Semantic Web Best Practices and Deployment Group* (2004-2005), *SKOS* (depuis 2005).

Mondeca se définit avant tout comme éditeur de logiciel. De 2000 à 2004 Mondeca s'est consacré à la conception et réalisation de son logiciel ITM (qui gère des ontologies, des bases sémantiques, des référentiels, des bases de connaissance) tout en participant à des projets de recherche européens. A partir de 2004, Mondeca a réalisé les premières installations opérationnelles de sa solution tout en affinant les fonctionnalités très génériques du logiciel pour répondre à des besoins clients (éditeur juridique, industrie automobile, tourisme...). Mondeca dispose, au moment de l'achèvement de notre travail de recherche, d'une offre complète de gestion de taxonomies et d'ontologies. Cette société est considérée comme un acteur de référence sur un marché en émergence. Mondeca emploie dix-sept personnes (juin 2008) et est organisé en quatre groupes : trois commerciaux (France, Europe du Nord, Etats-Unis), trois consultants avant-vente et installation client, un consultant très spécialisé sur les normes et les ontologies (l'« expert ontologue ») qui travaille sur pratiquement tous les projets, six développeurs

informatiques dont un « senior » spécialiste de l'ingénierie des connaissances, et un groupe de R&D « pure » (quatre personnes post-docs et doctorants) dirigé par Florence Amardeilh. Cette dernière a été recrutée en novembre 2004, pour réaliser une thèse en informatique, sur le couplage entre les outils d'extraction d'information et de représentation des connaissances (base de connaissance), à l'occasion du projet PressPro.

## **2. LES OUTILS DE L'ANNOTATION SEMANTIQUE ET DE L'EXTRACTION TERMINOLOGIQUE :**

### **2.1 Le Web Sémantique :**

Les logiciels que nous étudions s'inscrivent dans le cadre de ce qu'il est convenu d'appeler le Web Sémantique. Nous revenons ici brièvement sur ce concept, et plaçons en annexe des explications plus précises relatives aux notions d'annotation sémantique, d'extraction terminologique et de Ressources Terminologiques ou Ontologiques (RTO), qui contribuent à le fonder (cf. Annexe : « Web sémantique et notions liées »). L'objectif du Web Sémantique initié en 1998 par Tim Berners-Lee (1998), est de structurer les informations disponibles sur le Web. Pour cela, les ressources, textuelles ou multimédias, doivent être sémantiquement étiquetées par des métadonnées afin que les agents logiciels puissent les exploiter. La représentation explicite des contenus des ressources documentaires du Web est rendue possible grâce notamment aux ontologies. Ces ontologies (cf. Annexe) représentent une technologie clef pour la mise en œuvre de ce Web Sémantique, elles ont été développées en Intelligence Artificielle pour faciliter le partage de la connaissance et leur réutilisation. Depuis les années 90, les ontologies sont au cœur des recherches de différentes communautés, dont celles de l'ingénierie de la connaissance, du traitement automatique du langage naturel, et de la recherche d'information, les systèmes collaboratifs, etc. Cette popularité résulte en partie de ce que les ontologies proposent une compréhension commune et partagée d'un domaine, tant au niveau des utilisateurs humains qu'au niveau des applications logicielles. Parmi les autres propositions du Web Sémantique figurent la traçabilité de l'information, à savoir de sa provenance et de ses émetteurs afin de lui accorder un niveau de confiance adéquat, l'interconnexion des informations entre elles, la découverte de nouvelles connaissances à partir de raisonnements logiques.

La mise en place d'un Web Sémantique est susceptible d'apporter trois améliorations d'envergure au Web actuel, qui concernent les systèmes d'information des entreprises. La première est

L'amélioration des moteurs de recherche d'information par la capacité à effectuer des requêtes qui exploitent la structure de l'ontologie et qui infèrent de la connaissance à partir des informations existantes dans cette ontologie. La deuxième consiste à donner des moyens d'accès à l'information, permettant aux utilisateurs d'exploiter la connaissance représentée dans leur application. De tels moyens incluent des fonctionnalités pour trouver, partager, résumer, visualiser, naviguer et organiser la connaissance. La troisième est l'interopérabilité entre les différents systèmes d'information, particulièrement importante pour les entreprises souhaitant communiquer avec les réseaux de leurs prestataires par exemple. Dans ces cas-là, l'annotation basée sur l'utilisation d'une ontologie commune peut fournir un cadre de travail commun pour l'intégration d'informations provenant de sources hétérogènes. Pour ce faire, le Web Sémantique fournit un ensemble de langages et de technologies pour la modélisation des ontologies et l'annotation sémantique des contenus documentaires en fonction de ces ontologies (Uren et al., 2006).

## **2.2 Constats issus d'un pré-projet en 2002 :**

La réflexion et le travail effectués par Mondeca dans le cadre du projet destiné à un laboratoire pharmaceutique, en 2002, a débouché sur un certain nombre de constats, concernant les outils d'annotation sémantique. Un outil d'annotation sémantique est un logiciel qui permet d'insérer et de gérer des annotations sémantiques liées à une ressource documentaire donnée. La plupart des outils d'annotation sémantique ont évolué vers des environnements de plus en plus automatisés grâce aux méthodes issues des domaines de l'Extraction d'Information et des Systèmes d'Apprentissage. Il existait, en 2002, une gamme assez importante d'outils d'annotation sémantique, généralement issus de la recherche, de plus en plus orientés vers l'assistance des annotateurs humains à la création d'annotations. Cette réflexion a conduit Mondeca, à conclure que malgré une évolution rapide des langages et standards, ces outils n'étaient toujours pas adaptés à des utilisations concrètes en entreprise, du fait de certaines limites (Amardeilh, 2007) :

**1. Liens entre moteurs d'extraction et outils d'annotation sémantique :** les quelques outils d'annotation qui existaient, étaient intrinsèquement liés à un moteur d'extraction particulier, ce moteur étant complètement intégré à l'outil. Or, il peut être utile de recourir à différents moteurs d'extraction, suivant les besoins de l'application dans laquelle il s'inscrit. Il fallait donc travailler à la conception de solutions permettant de dissocier les deux logiciels pour que l'outil d'annotation puisse utiliser tel ou tel moteur d'extraction.

**2. Limites de l'apprentissage supervisé :** les moteurs d'extraction reposaient pour la plupart sur des processus d'apprentissage supervisés, systèmes performants avant tout pour du contenu structuré, mais pas pour du contenu non structuré qui recèle des informations telles que les relations sémantiques entre entités, contenu dont la sémantique est la plus difficile à extraire. Dans le cas des contenus non structurés, il est nécessaire de donner la priorité aux systèmes d'extraction basés sur des analyses sémantiques précises, donc sur des patrons d'extraction adaptés à un domaine particulier (cf. Annexe).

**3. Assistance aux utilisateurs pour le peuplement d'ontologie :** les processus et interfaces de ces outils n'offraient pas de suggestions ni de propositions aux utilisateurs, et ne prenaient pas en compte les contraintes modélisées dans l'ontologie de référence pour les guider.

A partir de ces constats, Mondeca et Temis ont cherché, plus tard, au travers des projets PressPro et Exinis, à construire une solution d'annotation sémantique, qui puisse aussi être utilisée pour peupler une ontologie. Le but était de réaliser un outil permettant de résoudre des cas concrets d'applications en entreprise, outil basé sur des outils du Traitement Automatique des Langues (TAL) et de Représentation des Connaissances. Il fallait pour cela définir une passerelle entre ces deux catégories d'outils, ce qui impliquait la résolution de certains problèmes, mis en évidence par la réflexion conduite précédemment :

**1. Problème de format :** comment passer d'un format de représentation du document textuel à un autre format de représentation dépendant de l'ontologie et de l'implémentation de la base de connaissance ? A l'issue de l'analyse linguistique, les outils de TAL produisent un ensemble d'étiquettes sémantiques le plus souvent sous la forme d'un document XML, aussi appelé arbre conceptuel. Alors que, de son côté, l'ontologie est modélisée dans le but de stocker et d'exploiter de la connaissance de manière rigoureuse et contraignante grâce aux formalismes de représentation des connaissances (tels OWL, RDF, XTM).

**2. Problème de la couverture du domaine** lors du passage du langage naturel au modèle : comment gérer le décalage entre le domaine couvert par l'ontologie et le vocabulaire contenu dans les ressources linguistiques pour les outils de TAL ? Les spécifications et le développement des ontologies et des ressources linguistiques utilisées par les outils de TAL sont indépendants les uns des autres. La couverture du domaine concerné n'est donc pas forcément alignée entre les étiquettes linguistiques et les concepts de l'ontologie, comme nous le verrons plus loin (p. 212).

En effet, l'ontologie peut servir à d'autres tâches impliquant une couverture du domaine différente de celle couverte par les outils du TAL. Ces derniers peuvent, quant à eux, réutiliser certaines ressources linguistiques existantes ou définir de nouveaux patrons linguistiques spécifiques au domaine étudié. Bien que justifié, le non-alignement de la couverture peut poser des difficultés d'intégration entre les deux systèmes, outre les contraintes techniques posées par l'implémentation même de la passerelle entre les deux types d'outils.

### 3. Problème de la conceptualisation lors du passage du langage naturel au modèle :

Comment savoir si la signification d'une étiquette sémantique produite par les outils du TAL correspond à la signification d'un concept de l'ontologie ? En effet, une étiquette sémantique a priori « identique » (ayant au moins le même libellé) à un concept de l'ontologie peut servir à annoter ou à instancier un tout autre concept de l'ontologie. Ceci dépend des besoins de l'application et de la modélisation de l'ontologie. A l'inverse, deux concepts de l'ontologie différents peuvent être instanciés à partir de la même étiquette sémantique et deux étiquettes sémantiques différentes peuvent servir à annoter ou à instancier le même concept de l'ontologie. Qu'est-ce qui peut servir à assimiler ou à distinguer une étiquette sémantique d'un concept de l'ontologie ?

Ces différents problèmes illustrent le fossé qui existait entre les outils d'extraction d'information et de représentation des connaissances (figure 6) :

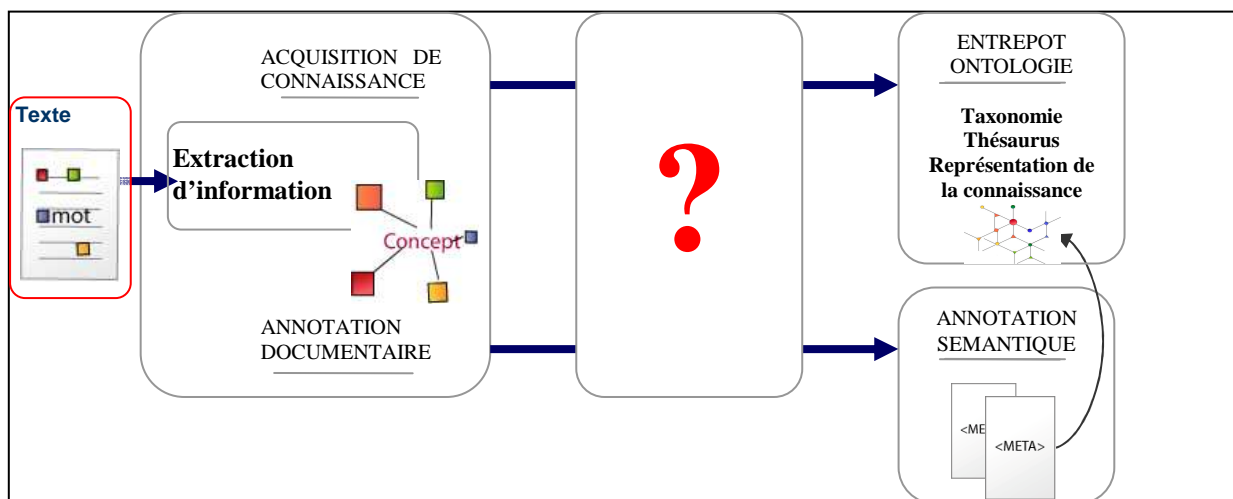


Figure 6 : Fossé entre la représentation textuelle et la représentation sémantique (d'après Amardeilh, 2007).

### 2.3 Un outil d'extraction d'information : *Insight Discoverer Extractor™* de Temis.

Le logiciel Insight Discoverer™ Extractor (IDE) est un logiciel destinés à l'extraction d'information (EI) commercialisé par la société Temis depuis juin 2002. Il constituait la brique la plus innovante du système que nous avons étudié, car l'extraction d'information était en 2003, et reste à ce jour, encore peu utilisée en entreprise, par rapport aux systèmes de représentation de connaissance, par exemple. Il était tout de même utilisé dans des projets industriels, notamment de veille économique et scientifique, pour de grands comptes français et internationaux. Son approche était fondée sur l'acquisition de connaissances à partir d'un corpus de textes selon un processus itératif, et exploitait les complémentarités des différentes étapes de l'analyse linguistique (morphologique, syntaxique et sémantique) à partir des étiquettes générées à chacune de ces étapes. IDE pouvait être employé pour une simple extraction d'éléments syntaxiques comme les noms, les verbes, etc. ou bien pour une extraction d'éléments sémantiques tels que des noms de sociétés, des noms de lieux, des dates, des prix,... et des relations sémantiques (fusion de X avec Y, achat de W par Z). L'architecture d'IDE reposait sur le moteur XeLDA™, issu des recherches du XRCE, un étiqueteur grammatical qui découpe chaque texte en unités lexicales, puis lemmatise ces unités lexicales pour qu'elles soient reconnues indépendamment de leurs formes fléchies (i.e. ramène les formes fléchies, conjuguées ou plurielles, à des formes standard, infinitif ou singulier) ; enfin, il assigne à ces unités lexicales une catégorie grammaticale (nom, adjectif, verbe...) assortie de traits morphosyntaxiques (genre, nombre). Pour réaliser cette analyse, le moteur XeLDA™ dispose de diverses ressources comme des dictionnaires, des règles morphologiques ainsi que des modèles statistiques pour résoudre les ambiguïtés concernant l'affectation des catégories grammaticales. Ces sources ne pouvaient être modifiées par l'utilisateur, elles étaient donc identiques quel que soit le domaine concerné.

Ensuite, à ce traitement, s'ajoutait l'analyse sémantique pour laquelle les ingénieurs linguistes de la société Temis définissaient des cartouches linguistiques, les Skill Cartridges™, en fonction du domaine concerné et des attentes de l'application finale. D'après Grivel et al. (2001), une Skill Cartridge™ est une hiérarchie de « *composants de connaissance* », appelés « Skill Units », décrivant l'information à extraire pour un métier, une activité, un domaine spécifique ou une thématique donnée. Un composant de connaissance peut avoir la forme d'un dictionnaire ou d'un ensemble de règles d'extraction qui décrivent l'information à extraire. Les règles d'extraction se composent d'un ensemble de patrons d'extraction contextuels, combinant lemmes, étiquettes syntaxiques et

étiquettes sémantiques. Chaque règle associe ensuite une nouvelle étiquette au fragment de texte repéré. Cette étiquette peut ensuite être utilisée dans de nouvelles règles.

Les linguistes de Temis suivaient les étapes de construction suivantes pour tout développement d'une nouvelle cartouche linguistique (appelée par Temis : Skill Cartridge™) :

1. L'analyse morphosyntaxique des textes qui constituent le corpus d'entraînement et la définition du vocabulaire pertinent relatif au secteur d'activité ou au domaine économique étudié (agroalimentaire, automobile, etc.) ;
2. Le regroupement des termes ainsi définis sous des étiquettes sémantiques elles-mêmes organisées, selon les besoins, en une hiérarchie de 3 à 4 niveaux généralement ;
3. La définition des règles d'extraction d'information niveau par niveau hiérarchique afin d'aboutir aux concepts visés : les acteurs du domaine d'activité étudié, leurs relations et toute information relative aux lieux, au temps ou à une donnée particulière, ou attribut, liée au domaine comme un montant financier par exemple ;
4. L'exécution interactive des règles d'extraction sur le corpus d'entraînement afin d'évaluer le résultat des extractions. Il était alors possible d'enrichir le vocabulaire, de modifier des règles d'extraction et d'en vérifier l'impact sur le corpus de travail.

En conclusion, IDE était un outil capable de s'adapter au domaine d'activité de l'entreprise cliente. Ces applications étaient développées par la propre équipe d'ingénieurs linguistes de Temis. Néanmoins, Temis et l'entreprise cliente devaient réaliser ensemble une analyse linguistique sur laquelle reposaient les patrons d'extraction construits manuellement. Au final, les informations étaient extraites et annotées par des étiquettes sémantiques constituant un arbre conceptuel. Nous allons maintenant étudier plus en détail la structuration des arbres de concepts générés afin de comprendre comment ils pouvaient être transformés en une représentation plus formelle du contenu.

### **La représentation en arbre conceptuel :**

La figure 7 (page 149) montre un exemple d'arbre conceptuel généré par IDE, à partir d'un extrait de l'article relatif à la biographie de « Francis Ford Coppola » présenté en Annexe.

La racine de cet arbre représente généralement le document analysé par le moteur d'extraction. Chaque nœud de l'arbre est constitué d'une étiquette sémantique (préfixée par le symbole « / ») et de la valeur (indiquée entre parenthèses) de l'unité textuelle du texte à laquelle cette étiquette a été affectée. D'autres informations, telles que le lemme de cette unité textuelle ou sa position dans le texte, peuvent aussi être associées aux étiquettes sémantiques si besoin est. Dans la Figure 7, la racine du document porte l'étiquette sémantique « /article » et a pour valeur son titre : «Famille Coppola, l'esprit de clan». Chaque sous-arbre correspond à une entité nommée isolée, à une phrase ou encore plus généralement à une proposition du document analysé.

Premièrement, les sous-arbres représentant les entités nommées isolées sont généralement formés à partir de deux schémas d'extraction :

1. Les entités nommées identifiées de manière « certaine » à partir d'un lexique ou de toute autre ressource linguistique utilisée par le moteur d'extraction ;
2. Les entités nommées identifiées de manière « potentielle » à partir de l'application d'un patron d'extraction compilé dans le moteur d'extraction.

Dans le premier sous-arbre de la figure 7, l'étiquette « /REFERENCE-ACTEUR » représente l'entité nommée isolée dont la valeur est « Francis Ford Coppola ». Ce nœud possède un nœud fils « /ActorNamed », qui signifie pour IDE qu'il a été repéré à partir d'un lexique, lui-même ayant pour nœud fils « /Personnalité » qui indique le lexique dont a été extraite l'entité nommée « Francis Ford Coppola ». A l'inverse, dans le deuxième sous-arbre de la figure 7, si l'entité nommée « Spike Jonze » a pour racine la même étiquette que l'entité nommée précédente, son nœud fils « /NomDePersonnePotentiel » indique que son étiquette sémantique a été calculée à partir du déclenchement d'un certain patron d'extraction pour la reconnaissance des noms de personnes construit à partir des étiquettes « Prénom » et « ProperName ». Le mélange des étiquettes ayant un label français ou anglais n'est pas un hasard. Les ingénieurs linguistes réutilisent souvent des ressources linguistiques d'applications différentes pour éviter de réécrire tous les patrons à chaque nouvelle application. Dans notre cas, l'étiquette « ProperName » avait été définie dans une précédente cartouche linguistique.

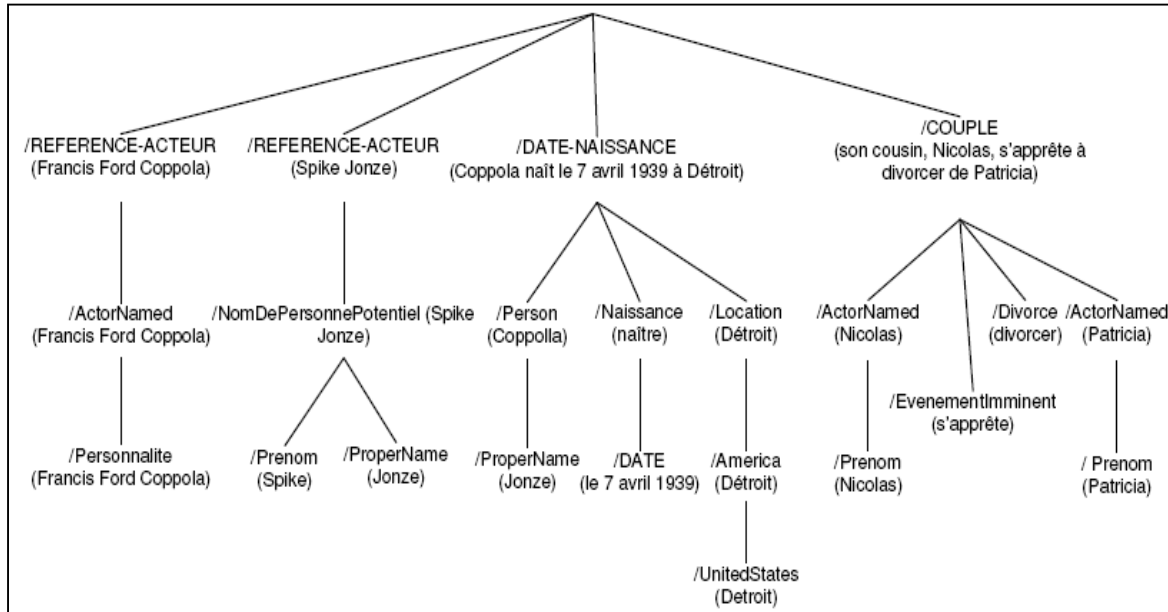


Figure 7 : Extrait de l'arbre conceptuel produit par l'analyse linguistique de l'article « la tribu Coppola » publié dans le magazine Elle.

Deuxièmement, une proposition n'est représentée sous la forme d'un sous-arbre que si elle est composée d'unités textuelles ayant donné lieu à une extraction. De la même manière, certaines unités textuelles de la proposition peuvent ne pas apparaître dans le sous-arbre : seules celles ayant été étiquetées sémantiquement par les patrons seront représentées par un nœud du sous-arbre. Ces sous-arbres correspondent à l'interprétation de la proposition pour répondre aux tâches de reconnaissance d'un attribut, d'une relation, voire d'un scénario. D'après l'analyse des sous-arbres, on peut résumer les différentes situations comme suit : lorsqu'ils contiennent une seule référence à une entité nommée, alors la tâche correspond à l'identification d'un attribut pour celle-ci ; lorsqu'ils contiennent deux étiquettes correspondant à des entités nommées, alors il s'agit de la tâche d'instanciation d'une relation entre ces entités nommées ; lorsqu'ils contiennent plusieurs entités nommées, cela signifie qu'un scénario a été clairement identifié au sein de la proposition.

Par exemple, le sous-arbre portant l'étiquette « /DATE-NAISSANCE » dans la Figure 8, ci-après, correspond à l'extraction d'un scénario complet correspondant à la naissance d'une personnalité. Ce scénario est constitué des entités nommées « Personnalité », « Date » et « Lieu », respectivement représentées par les étiquettes « /Person », « DATE » et « Location ».

Chaque proposition, ou sous-arbre, constitue un fait qui représente la structure schématique d'un événement, d'une action, d'un état, etc. Ainsi, chaque phrase peut être représentée en distinguant la proposition et ses modalités. La proposition est constituée d'un verbe d'où partent des relations étiquetées sémantiquement et correspondant aux différents « cas ». Les modalités concernent les informations qui situent et spécifient la proposition comme la négation, le mode, le temps, l'aspect, etc. Il suppose que pour chaque verbe, il existe seulement un nombre restreint de cas parmi les suivants :

1. AGENT : l'instigateur animé d'une action
2. INSTRUMENT : la force inanimée ou l'objet affecté
3. DATIF : l'animé affecté par l'action
4. FACTITIF : l'objet résultat de l'action
5. LIEU : le lieu ou l'orientation
6. OBJET : l'entité qui bouge, change ou dont la position ou l'existence est en question.

La relation entre grammaire des cas et arbre conceptuel est encore plus évidente dans le domaine de la veille économique, bien connu des moteurs d'extraction d'information, et d'IDE en particulier. Les scénarios de ce domaine sont généralement bien identifiés et structurés en fonction des différentes informations à extraire correspondant aux questions suivantes : Qui sont les acteurs du secteur d'activité ou du domaine économique étudié ? Quels sont les objets considérés relatifs au domaine décrit ? Quelles sont les actions de ces acteurs, sur quels objets portent-elles, comment s'effectuent-elles ? Où ont lieu les actions en question ? Quand ont-elles eu lieu ? Quel est le montant de ces actions ?

Mais cette grammaire des cas ne peut correctement représenter les phrases plus complexes où, comme-ci-dessus, on a des emboîtements de propositions les unes dans les autres. Elle convient surtout aux phrases dites simples. L'arbre conceptuel est capable de représenter hiérarchiquement des constructions plus complexes comme par exemple dans le dernier sous-arbre de la figure 8 où le syntagme nominal « son cousin, Nicolas » se décompose en deux syntagmes nominaux : le

premier, « son cousin », est représenté par l'étiquette sémantique « /Famille » et le second « Nicolas » identifié par l'étiquette « /Prénom ».

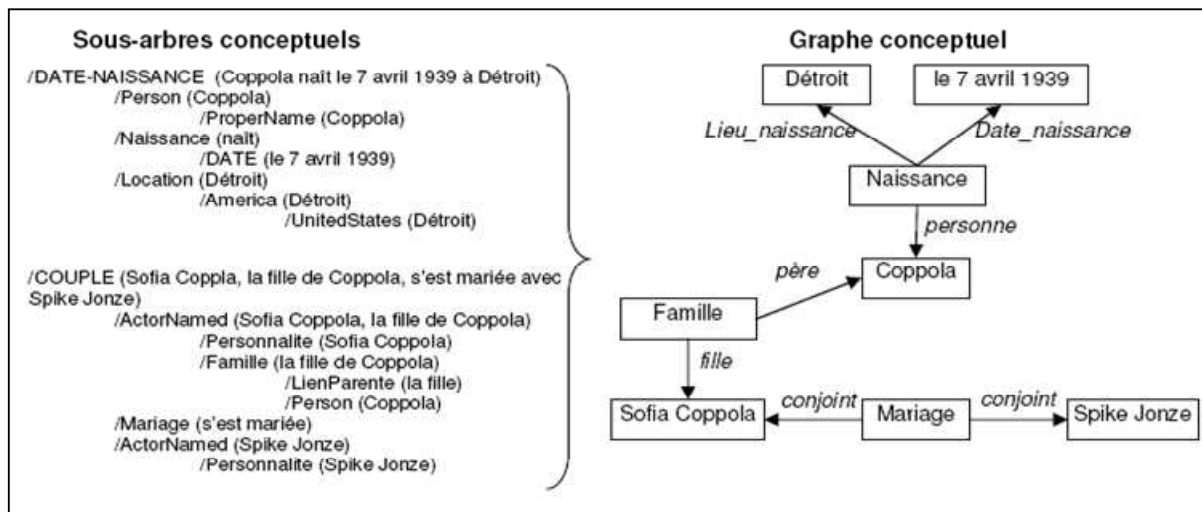


Figure 8 : Transformation de deux sous-arbres conceptuels en un graphe conceptuel où la même entité nommée « Coppola » fait le lien entre les deux propositions. (Amardeilh, 2007).

Pour qu'un arbre conceptuel issu de l'analyse linguistique produise une représentation plus globale de la signification du texte original, il faut pouvoir identifier et référencer les entités nommées identiques entre elles au delà des différents sous-arbres des propositions. Comme illustré à la figure 8, nous obtiendrons alors une représentation sous la forme d'un graphe conceptuel qui est un premier pas vers une représentation ontologique.

## 2.4 Synthèse et ouverture sur l'innovation à développer :

Pour développer une solution automatisée d'annotation sémantique et de peuplement d'ontologie, il fallait parvenir à faire correspondre les différentes étiquettes sémantiques de l'arbre conceptuel avec les éléments (concepts, relations et attributs) de l'ontologie de référence. Les étiquettes étaient habituellement créées par les ingénieurs linguistes lors du développement des patrons d'extraction alors que les éléments de l'ontologie préexistaient ou étaient modélisés par un expert en représentation des connaissances et en ingénierie ontologique, appelé « ontologue ». Par conséquent, il était nécessaire d'interpréter correctement la sémantique fournie par les arbres conceptuels mais également de prendre en considération le décalage qui pouvait exister entre les deux modes de représentation de la connaissance.

Dans la suite, nous allons exposer la solution développée pour et grâce aux projets PressPro et Exinis France, principalement par F. Amardeilh (2007), responsable chez Mondeca du couplage entre IDE et ITM, en collaboration avec l'expert ontologue de Mondeca et avec les linguistes de Temis : cette solution repose sur la définition d'un ensemble de règles nommées « Règles d'Acquisition de Connaissance ». Ces règles permettent notamment de coupler une ou plusieurs étiquettes sémantiques d'un arbre conceptuel avec un élément de l'ontologie de référence. Elles s'appuient sur le contexte des étiquettes sémantiques dans l'arbre conceptuel, ce afin de résoudre un certain nombre d'ambiguïtés.

Dans le processus d'annotation sémantique, les ontologies jouent un rôle primordial puisqu'elles modélisent les concepts, leurs attributs et les relations utilisées pour annoter le contenu des documents. L'ontologie contraint l'application sur les vocabulaires et les instances autorisés comme métadonnées. Mais s'il est essentiel pour une application Web Sémantique de reposer sur une ontologie pour la réalisation de cette tâche d'annotation sémantique, il est aussi important que la base de connaissance, associée à cette ontologie, contienne les instances à utiliser pour l'annotation sémantique. C'est pourquoi la tâche de peuplement d'ontologie a pour but d'enrichir semi-automatiquement la base de connaissance avec de nouvelles instances de concepts, d'attributs et de relations comme défini par la modélisation de l'ontologie.

En fait, ces deux tâches sont proches. Premièrement elles reposent toutes deux sur la modélisation de ressources terminologiques et ontologiques (comme les taxonomies, thésaurus, ontologies) pour normaliser la sémantique des annotations documentaires comme celle des concepts du domaine concerné. Deuxièmement, elles utilisent les méthodes et outils du Traitement Automatique du Langage Naturel (TALN), comme l'Extraction d'Information (EI), pour extraire une information structurée des ressources documentaires ou encore la Catégorisation pour classifier un document dans des catégories prédéfinies ou calculées. Troisièmement, elles s'appuient de plus en plus sur les standards et langages du Web Sémantique comme RDF pour l'annotation et OWL pour le peuplement.

L'innovation qui a été développée au travers des projets PressPro et Exinis consiste fondamentalement à combiner des outils d'extraction d'information et des outils de représentation des connaissances pour la réalisation de ces deux tâches. La solution ainsi conçue intègre ces différents outils tout en permettant le remplacement du moteur IDE par un autre, mais aussi le remplacement du système de représentation des connaissances ITM par un outil

équivalent. Cette combinaison constitue en soi une innovation, les rares solutions existantes étant encore généralement au stade de la recherche, et intégrant l'outil d'extraction d'information à la base de connaissance, sans possibilité de découplage. Cette solution innovante regroupe différentes briques logicielles, dont certaines constituent elles-mêmes des innovations, en l'occurrence le moteur d'extraction IDE (reposant sur une technologie encore peu utilisée en entreprise) et la passerelle logicielle OntoPop reliant IDE à ITM.

### **3. LE PROJET PRESSPRO :**

#### **3.1 L'origine du projet :**

En 2002, la société Novartis avait donc incité Mondeca à s'intéresser au peuplement automatique d'ontologie, et pour cela à travailler avec Temis. Ensuite, Mondeca a passé, début 2003, un accord OEM avec Temis. Cela signifiait que Mondeca pouvait théoriquement proposer à ses clients, d'intégrer les fonctionnalités d'extraction d'information du logiciel IDE de Temis à son système de représentation des connaissances ITM. Pour Novartis, un prototype avait été développé, ainsi qu'une passerelle logicielle entre les logiciels IDE de Temis et ITM de Mondeca : mais cette passerelle était difficilement modifiable, alors que cela était nécessaire dès lors que les arbres de concept générés par l'outil d'extraction d'information IDE évoluaient. Le couplage tel qu'il était réalisé, impliquait donc le développement d'une passerelle complètement nouvelle à chaque projet. En cette occasion, Temis et Mondeca se sont aussi rendus compte de leur complémentarité, car Mondeca souhaitait enrichir ses ontologies en récupérant l'information extraite par le moteur d'extraction de Temis, tandis que Temis voyait au travers du peuplement d'ontologie une application nouvelle pour son logiciel IDE.

Par la suite, en 2003, Temis, Mondeca et Xyleme ont participé au développement d'une offre logicielle, visant à doter le service documentation du groupe PressPro, d'une base de connaissance relative aux « people » ou « célébrités ». Cette base de connaissance devait permettre aux documentalistes du groupe PressPro de constituer, plus efficacement et plus rapidement, des dossiers sur ces « célébrités », à l'intention des journalistes du groupe. Ainsi, Temis, Mondeca et Xyleme ont travaillé à ce projet d'octobre 2003 à novembre 2006. Ces trois éditeurs ont dû faire évoluer chacun de leurs logiciels afin de les adapter aux besoins de PressPro, et collaborer entre eux pour réunir leurs logiciels respectifs dans une seule et même application.

Temis et Mondeca ont ainsi dû travailler au développement d'une brique logicielle intermédiaire entre leurs logiciels IDE et ITM. Cela leur a demandé d'innover à plusieurs niveaux :

- la brique logicielle IDE, moteur d'extraction d'information, constituait une innovation en ce que cette technologie était encore peu mature et peu répandue, par rapport à celle des systèmes de gestion de base de connaissance ou du stockage de données au format XML,
- le couplage entre outils d'extraction d'information et de représentation de connaissance, dans le cas d'outils initialement indépendants l'un de l'autre, n'avait pas encore été réalisé pour des applications commerciales. Il n'existait pas de cas répertorié de solution d'automatisation d'annotation sémantique et de peuplement d'ontologie dans le domaine de la presse et de l'édition.

## 3.2 Description et chronologie du projet

### 3.2.1 *La solution à mettre en place :*

Le projet concernait la presse « People », un secteur en plein essor. Lorsque les journalistes écrivaient un nouvel article pour une publication du groupe PressPro, ils demandaient aux documentalistes de leur fournir des informations au travers de fiches de synthèse ou d'articles déjà publiés. Ces fiches et articles étaient stockés dans une base de données, depuis une dizaine d'années. Cette base était perçue comme trop lourde et trop difficile à interroger par les documentalistes du groupe, les interfaces de recherche étant particulièrement complexes.

Le système en place, jusqu'en 2003, reposait sur :

1. La lecture quotidienne de la presse par les documentalistes ;
2. L'indexation et le classement manuels des articles intéressants à l'aide de divers thésaurus et autres vocabulaires contrôlés ;
3. La mise à jour manuelle de la base de données recueillant les fiches signalétiques.

Ce système ne proposait donc pas d'indexation automatique, ce qui contraignait les documentalistes à consacrer 80% de leur temps à indexer manuellement les articles de presse, pour pouvoir les archiver. Ainsi PressPro a ressenti en 2003, le besoin de faire évoluer cette base

de données, et de bénéficier des possibilités offertes par les systèmes de représentation des connaissances. L'objectif du projet était donc d'aider les documentalistes en automatisant tout ou partie des trois tâches évoquées, mais aussi en facilitant la formulation de requêtes, afin qu'elles puissent répondre plus rapidement et plus précisément aux demandes des journalistes.

La nouvelle application, quant à elle, devait permettre aux utilisateurs de travailler de la manière suivante :

*« Les documentalistes de PressPro sélectionnent un article de presse, qu'elles souhaitent archiver. Cet article est alors traité par le logiciel de Temis, qui le classe dans une rubrique du plan de classement de PressPro. Ensuite, cet article est indexé, c'est-à-dire que le logiciel extrait semi-automatiquement des informations (i.e. il propose des extractions que la documentaliste doit valider). Certaines extractions servent à annoter le document (comme par exemple, la liste des personnes citées dans le document), et sont envoyées dans le logiciel Xyleme, qui réalise l'archivage XML des articles. Ce document XML annoté est placé dans une catégorie (spectacle, sport, etc...). D'autres extractions, qui concernent les relations entre gens célèbres (exemple : « mariage ») sont envoyées dans la base de connaissance de Mondeca (« peuplement de l'ontologie »).*

*Ensuite, quand un journaliste du groupe PressPro, demandera un dossier sur un thème (exemple : toutes les informations concernant les mariages des chanteurs depuis 1998), les documentalistes pourront faire appel (via l'application documentaire) soit au logiciel Xyleme (pour retrouver des articles sur le chanteur), soit au logiciel Mondeca (pour des informations relatives aux mariages des chanteurs, pour éventuellement retrouver ensuite dans le logiciel Xyleme, les articles dont cette connaissance est extraite). »*

### 3.2.2 La chronologie du projet :

Maintenant que nous avons décrit le fonctionnement du produit, nous voudrions revenir sur le processus qui a permis de le façonner :

1. Premier contact en juin 2003, entre PressPro, Mondeca, Temis et Xyleme.
2. Rédaction d'une première proposition par Temis et Mondeca en septembre 2003, refusée par PressPro.
3. Rédaction d'une deuxième proposition incluant Xyleme, acceptée en octobre 2003 : signature d'un premier contrat.

4. Remise d'un document de spécifications à PressPro en mars 2004, et signature du contrat définitif en mai 2004.
5. Ateliers de travail avec des *user groups* ou groupes d'utilisateurs, de mai à décembre 2004.
6. Livraison d'une version V0 du système et Arrêt des *user groups* en décembre 2004.
7. Transfert de la responsabilité de l'application documentaire de PressPro à Xyleme en février 2005.
8. Version V1 (production) Beta livrée en juin 2005, sans la « brique » de Temis.
9. Test de la version V1 Beta de juin à octobre 2005.
10. Mise en production du système (V1) en février 2006, sans la « cartouche » de Temis.
11. Fin du consortium en mars 2006, et départ de la chef de documentation de PressPro.
12. Test de la brique de Temis, et rejet par les documentalistes en juillet 2006.
13. Arrêt définitif du système à partir de novembre 2006.

### **3.3 Les phases amont jusqu'à la signature du contrat définitif :**

Nous présentons les quatre premières étapes ayant conduit à la signature du contrat définitif en mai 2004.

#### *3.3.1 Première étape avant la signature d'un premier contrat :*

##### **Le premier contact en juin 2003 :**

Le premier contact relatif au projet étudié, a eu lieu, au salon i-expo (salon de l'informatique numérique se déroulant à Paris) en juin 2003. Temis a reçu la visite de Catherine, la chef de documentation de PressPro, et de la responsable de la formation de cette entreprise. Sur ce même salon, ces personnes avaient déjà rencontré Mondeca. Les contacts furent ainsi établis entre la chef de documentation et le Vice President Sales Marketing, de Temis, et le Directeur des Ventes, de Mondeca. Rapidement, l'idée est venue à la chef de documentation de demander à Temis et Mondeca de faire une première proposition conjointe, un « draft » donnant une estimation du coût d'un nouveau système documentaire permettant l'indexation automatique de documents, et la création d'une nouvelle base de connaissance. A l'époque, la relation entre Temis et Mondeca

était déjà structurée car un contrat OEM avait été signé entre Temis et Mondeca, ce que savait PressPro. Cela signifie que Mondeca pouvait commercialiser des solutions, intégrant à l'intérieur de ses propres logiciels des briques logicielles fournies par Temis, et ce de manière généralement transparente pour l'utilisateur. Néanmoins, en juin 2003, cette possibilité ne s'était pas encore concrétisée car aucun client n'avait encore acheté de solution de ce type.

### **Le premier draft en septembre 2003 :**

Les responsables commerciaux de Temis et de Mondeca, ont rédigé une première proposition, qui reposait sur les deux idées suivantes : Mondeca dirigerait le projet (rédaction des spécifications, développement proprement dit et intégration chez PressPro), et serait donc le seul interlocuteur de PressPro ; tandis que Temis jouerait uniquement le rôle d'OEM, fournissant les logiciels dont Mondeca aurait besoin. Cette première tentative fut un échec, Mondeca s'estimant insuffisamment compétent pour, dans un premier temps, chiffrer correctement ce projet et dans un deuxième temps le conduire seul. Ensuite, les deux responsables commerciaux rédigèrent une proposition attribuant à Temis et Mondeca un rôle équivalent dans le projet, et le texte final fut achevé en septembre 2003. Leur proposition a été transmise en septembre 2003 à la chef de documentation, et au Directeur des Systèmes d'Information de PressPro (qui était le Directeur Financier, le poste de Directeur des Systèmes d'Information du groupe PressPro étant alors vacant).

La solution consistant à faire de Mondeca le chef de projet et le seul interlocuteur de PressPro n'était pas envisageable car le problème posé par PressPro au niveau de l'extraction de l'information était beaucoup trop complexe pour Mondeca seul, qui n'est pas un spécialiste du *text mining*. Temis et Mondeca se sont donc estimés contraints de se positionner, dans un deuxième temps, comme partenaires pour répondre à la demande de PressPro. Ils le firent alors qu'ils ne l'avaient pas prévu au départ, que cela ne correspondait pas à l'accord OEM qu'ils avaient signé précédemment, et qu'ils redoutaient la complexité d'une telle organisation. A ce moment-là, Xyleme, le troisième éditeur, ne faisait pas encore partie du projet.

### **Une deuxième proposition, incluant un troisième éditeur, Xyleme, imposé par le client :**

La négociation (en septembre et octobre 2003) s'est déroulée entre les responsables commerciaux déjà cités et la chef de documentation, Catherine. Pendant cette négociation, cette dernière a

suggéré à Temis et Mondeca, de travailler avec Xyleme, spécialiste de l'archivage XML, pour formuler une proposition à trois. Malgré leurs réticences, Temis et Mondeca ne se sont pas opposés de manière absolue à l'intégration de Xyleme dans le projet. En septembre et octobre 2003, il y eut une série de discussions et un cahier des charges fut rédigé : les éditeurs le nommèrent « cahier des charges fonctionnel » et PressPro, « schéma directeur du nouveau système documentaire ».

Xyleme avait mis en place une solution d'archivage au « Monde », et Catherine avait ainsi pu apprécier indirectement le travail de Xyleme. Par contre, Mondeca n'avait pas réellement de références dans la presse, mais PressPro a rapidement estimé que Mondeca était en mesure de proposer un système, à la fois évolutif et ouvert. PressPro ne voulait pas d'intégrateur (SSII) sur ce projet contrairement à Temis et aux deux autres éditeurs de logiciel, qui pensaient que cette configuration serait difficile à gérer, même si Xyleme faisait la coordination globale du projet.

### *3.3.2 Un premier contrat en octobre 2003 : la rédaction des spécifications du projet.*

#### **Les conditions de ce premier contrat en octobre 2003 :**

Dès le mois d'octobre 2003, les trois éditeurs ont donc travaillé ensemble sur le projet. Il y a eu un premier contrat de 60 k€ signé en octobre pour la rédaction des « spécifications » par les trois éditeurs : ce sont les responsables de l'avant-vente et les responsables du projet de, respectivement, Temis (deux personnes), Mondeca (deux personnes) et Xyleme (une personne) qui l'ont fait. Il s'agissait d'un premier contrat, d'une valeur limitée portant sur la rédaction des spécifications fonctionnelles (quelles fonctions doit posséder le logiciel ?) et des spécifications techniques (quels choix techniques retenir ?) du futur système. Il n'y a pas eu de problèmes particuliers de définition des rôles, entre les trois éditeurs de logiciels. Un chef de projet « officiel », appartenant à Xyleme, a été nommé pour le « consortium », car vis-à-vis de PressPro, Xyleme/Temis/Mondeca s'est positionné comme un consortium, bien que celui-ci n'eût en fait pas de statut juridique particulier, le terme ne constituant qu'une simple dénomination.

Fondamentalement, ce projet multi-acteurs était totalement nouveau tant pour le client que pour les trois membres du consortium. Les spécifications devaient donc être définies sans qu'il puisse être fait référence à des projets similaires, et en conséquence leur rédaction a constitué un processus plutôt exploratoire.

### **La rédaction des spécifications d'octobre 2003 à février 2004 :**

Des réunions se déroulèrent régulièrement, au moins une fois par semaine, chez PressPro pour comprendre le besoin du service « Documentation », avec Catherine et Fabrice qui était le responsable informatique de ce service. Ces réunions eurent lieu d'octobre 2003 à janvier 2004, chez PressPro avec les cinq personnes du « consortium ». Au début, Catherine et Fabrice ont présenté l'activité de PressPro, comment les utilisateurs (les documentalistes) travaillaient, avec quels référentiels, selon quels processus, ce qu'ils indexaient, où et comment ils stockaient ce qui était indexé, quelles étaient ensuite les fonctionnalités de leurs recherches. Il fallait que les trois éditeurs de logiciel sachent ce que les documentalistes voulaient constituer comme dossiers, comment elles souhaitaient les publier, par quels moyens elles voulaient retrouver l'information. Catherine avait la vision du métier de documentaliste la plus claire, alors que Fabrice se focalisait sur l'architecture informatique. C'était fondamentalement sur le métier de documentaliste chez PressPro, que le consortium avait besoin d'informations. Parallèlement, des réunions réunissant uniquement les trois éditeurs se déroulaient une à deux fois par semaine, d'octobre 2003 à février 2004. Lors de cette première phase, le consortium avait interdiction de contacter des utilisateurs (service documentation d'PressPro) car le projet était secret. En effet, un nouveau projet de ce type doit obligatoirement passer en conseil d'établissement où siègent les syndicats. Derrière ce projet, se profilait une réorganisation complète du département documentaire : la direction de PressPro n'avait pas vraiment prévu de plan social, mais une redéfinition des tâches. L'objectif était quand même de diminuer de 20 ou 30 % l'effectif du service documentaire, qui comptait environ 50 documentalistes, grâce à des mutations internes.

### **Un pré-rapport, le 15 décembre 2003 :**

Un pré-rapport concernant les spécifications a été remis à PressPro le 15 décembre 2003, et une présentation faite à Catherine pour montrer la bonne compréhension du problème par le consortium. C'était une étape de pré-validation des spécifications, le document n'était constitué que d'une vingtaine de transparents, pour montrer l'architecture envisagée. Catherine a immédiatement validé ce travail et demandé sa poursuite. Après la validation, du 15 décembre 2003, il a fallu rédiger le document. Les éditeurs du consortium estimaient que Catherine avait bien compris ce qu'il lui avait été présenté, bien qu'elle ne soit pas informaticienne de formation.

### **Une mise en concurrence éphémère en décembre 2003 et janvier 2004 :**

A ce moment-là, en décembre 2003, PressPro a voulu mettre un concurrent sur le projet, le groupe Ever, un acteur de la Geide (gestion électronique de document). Le patron de ce groupe bien implanté chez PressPro, a su qu'il y avait un projet en cours et il a fait le forcing auprès d'un des dirigeants de PressPro qu'il connaissait bien. Celui-ci a essayé de faire en sorte qu'Ever soit consulté. Catherine était, elle, persuadée que le « consortium » (terme utilisé par PressPro pour désigner le groupe d'éditeurs de logiciel impliqués dans le projet) répondait à sa problématique, alors qu'Ever n'était pas, à ses yeux, un spécialiste de l'extraction ni de la catégorisation de l'information. Toutefois, cela a contraint le « consortium » à faire une présentation le 20 décembre 2003, devant le Directeur Financier et DSI. Une quinzaine de transparents ont été présentés par les trois éditeurs, afin de montrer que la solution qu'ils préconisaient n'était pas comparable à ce que pouvait proposer Ever, en ce qu'elle était beaucoup plus « verticale ». Aussi au mois de janvier 2004, la poursuite du projet par le « consortium » a-t-elle été approuvée par la direction de PressPro.

### **La remise du document de spécifications à PressPro en mars 2004 :**

Ce travail s'est achevé fin février 2004, avec un mois de retard par rapport aux prévisions. Il a abouti à la remise d'un document de spécifications (plus de cent pages) à PressPro : il comportait le besoin du client, la solution proposée, la manière dont les trois éditeurs voyaient le système global, et puis la description des flux et la modélisation des processus. Le document comprenait donc l'architecture, et les diagrammes UML (description de ce que fait le logiciel tâche après tâche), mais il s'agissait de spécifications très fonctionnelles, et elles n'étaient pas très détaillées. Ce rapport a été analysé chez PressPro par Catherine et le DSI. Dans le document, il n'y avait pas de chiffrage des coûts, seulement les spécifications, avec un planning. Ensuite, l'hypothèque Ever étant levée, les choses évoluèrent rapidement, car dès mi-mars 2004, PressPro a demandé une cotation.

Ce travail était exploratoire pour les trois éditeurs, dans la mesure où il y avait plusieurs chantiers à spécifier, dépendants du métier de PressPro (que Mondeca et Temis ne connaissaient pas), et de ses archives. Pour Temis, il s'agissait de reprendre toutes les données existantes, de réaliser le développement spécifique d'une « cartouche de connaissance » (i.e. logiciel capable d'extraire d'articles quelconques, des informations pertinentes relatives à la vie des célébrités), la

catégorisation et l'indexation de ces articles. Pour Xyleme, qui parmi les trois éditeurs du consortium devait jouer le rôle de chef de projet logiciel, cela constituait aussi un gros chantier pour comprendre la problématique d'ensemble, les flux documentaires à absorber. Pour Mondeca, le travail consistait à rédiger les spécifications concernant la création de la base de connaissance, et la constitution de dossiers de recherche (dossiers que doivent bâtir les documentalistes, suite la demande d'un journaliste). Il y a aussi eu beaucoup de spécifications relatives aux interfaces de validation (des propositions d'indexation) par les documentalistes. Il fallait aussi développer des interfaces de recherche, pour la recherche d'information et la constitution de dossiers.

Ces spécifications ont pris beaucoup de temps aux différents acteurs, dont les linguistes de Temis, qui jugèrent ce travail très formateur parce que l'équipe de Xerox (Business Unit XTS/Xelda) venait d'intégrer Temis. Cela a permis aux équipes grenobloises (issues du XRCE et toujours située dans les locaux de Xerox) et parisienne (équipe originelle) de Temis, de mieux se connaître.

#### **Les garanties juridiques demandées par PressPro en mars 2004 :**

Les trois éditeurs ont dû rédiger un document, intégrant des garanties juridiques. Différents documents juridiques, ainsi que le contrat de licence des logiciels, ont été rédigés et échangés pendant plusieurs semaines. Les principaux points abordés concernent les garanties relatives à la disparition éventuelle d'un éditeur, ou son incapacité à fournir le travail attendu. En résumé, la philosophie de PressPro était simple : « si cela ne marche pas, vous remboursez », ce qui signifiait que chaque éditeur de logiciel était responsable à hauteur de sa contribution. Par contre, si la solution ne fonctionnait pas par la faute d'un seul éditeur, dans quelle mesure les deux autres devaient-ils rembourser ? ... Le problème était complexe pour Temis, Xyleme et Mondeca. Sur ce point, les négociations sont remontées à la direction de Temis, de Mondeca et de Xyleme, qui ont traité directement avec le juriste responsable de ce contrat chez PressPro, ce qui a pris du temps.

#### **Le chiffrage du projet en mars 2004 :**

La proposition définitive d'une quinzaine de pages a été remise début avril 2004, sans nouvelle présentation. Cette proposition a été traitée par la chef de documentation et le DSI. La

proposition faite par le consortium pour l'ensemble du projet, était de l'ordre de 650 k€, répartis de la manière suivante :

- 300k€ de licences de logiciel,
- 300k€ de développements,
- 45 k€ annuels pour la maintenance,
- auxquels s'ajoute la formation du personnel de PressPro facturée à part,
- et bien sûr les 60 k€ déjà facturés pour la rédaction des spécifications.

Au départ, les trois éditeurs avaient chiffré chacun de leur côté. La cotation n'avait alors plus rien à voir avec celle qui avait été présentée à PressPro dans le premier draft de l'été 2003, car elle atteignait une valeur très nettement supérieure. Le chiffrage a été présenté sous la forme d'une offre globale avec trois éditeurs, une licence par éditeur, et ensuite des « jours-hommes » pour le développement spécifique. La maintenance devait devenir effective à partir de la mise en production des logiciels (une fois la solution utilisée par le client).

Ensuite il y a eu une réunion entre, d'une part, la chef de documentation et le DSI, et, d'autre part, les responsables commerciaux de chaque éditeur. L'accord final est alors intervenu sur la base de ce proposaient les trois éditeurs à PressPro, moins 10%, soit 270k€ de licences de logiciel, 270 k€ de développement (sous forme de forfait), 40 k€ annuels pour la maintenance auxquels s'ajoutent la formation, et les 60 k€ déjà facturés pour les spécifications

### *3.3.3 Le contrat définitif entre les trois éditeurs signés en mai 2004 :*

Le contrat final portant sur le développement d'une solution complète d'archivage d'articles « people » destiné au service documentation de PressPro fut signé en mai 2004. Les principaux livrables prévus à la date de la signature étaient les suivants :

1. Des livraisons intermédiaires de logiciels devaient avoir lieu en novembre 2004 (préversion du logiciel d'extraction de Temis), en décembre 2004 (version quasi-définitive des logiciels de Xyleme et de Mondeca), et le système complet (logiciels de Temis, Xyleme et Mondeca plus l'application documentaire qui doit fédérer ces trois logiciels) devait être en pré-production chez PressPro en février 2005, c'est-à-dire qu'une version devait être mise en place pour réaliser des tests.

2. Le système devait être en production (utilisation complète par les documentalistes de PressPro) en juin 2005.

Entre le premier contact et la signature définitive du contrat de développement, se sont donc écoulés onze mois. En quatre mois (entre juin et octobre 2003), les éditeurs de logiciel retenus par PressPro ont eu une mission de spécifications à réaliser. La partie « spécifications » a été vécue par les trois éditeurs comme une phase pendant laquelle PressPro les a testés. La rapidité de cette première contractualisation, est venue de ce que PressPro ressentait un besoin fort de faire évoluer son système. Le choix fait par PressPro, de trois petits éditeurs de logiciels, venait de ce que ce client recherchait des spécialistes dans chaque domaine (extraction/catégorisation d'information pour Temis, archivage XML de documents pour Xyleme, base de connaissance pour Mondeca). Pour les responsables du projet du consortium, PressPro les avait retenus parce qu'ils étaient français (proximité culturelle et linguistique), proches physiquement (réunions et collaborations facilitées), et potentiellement réactifs, et ce en écho aux problèmes rencontrés avec OpenText. OpenText dont la solution avait été mise en place chez PressPro par un intégrateur français, et n'était pas intervenu (pas plus que l'intégrateur) de manière satisfaisante afin de la rendre opérationnelle.

Suite à la signature du contrat, une réunion a été organisée pour annoncer le projet aux documentalistes. Il y a eu un meeting mi-mai 2004, où les éditeurs ont présenté le projet aux 50 documentalistes, mais dans un climat tendu. Selon eux, cela résultait des difficultés rencontrées avec l'ancien système, mais aussi de ce que certains savaient que PressPro avait prévu des réductions d'effectif. La chef de documentation et le responsable de l'informatique documentaire, présents, ne sont pas intervenus au cours de cette réunion.

### **Les logiciels à développer (à partir de mai 2004) :**

La tâche de chaque éditeur de logiciel et de PressPro définie lors du lancement du projet, pouvait être résumée de la manière suivante :

1. Temis devait créer, d'une part, une « cartouche de connaissance » (couplée à IDE) permettant d'extraire les informations « people » de tout article de presse, et d'autre part adapter son logiciel IDK, module d'indexation et de catégorisation de ces articles.

2. Mondeca devait créer une nouvelle base de connaissance, après avoir défini l'ontologie du domaine, et concevoir la passerelle logicielle permettant de récupérer les informations extraites par IDE,

3. Xyleme devait récupérer les articles appartenant au fond documentaire de PressPro (deux millions d'articles), puis réaliser un module d'archivage XML des articles de presse, acceptant les articles traités par le logiciel de Temis,

4. PressPro devait développer, en collaborant avec les trois éditeurs, l'application documentaire fédérant les différents composants développés par les trois éditeurs, et qui constitue l'interface utilisateur du système, pour les documentalistes.

### 3.4 L'organisation du projet à partir de mai 2004 :

#### 3.4.1 Les acteurs du projet :

Nous reprenons ici le profil en termes de compétences et de rôle de chacun des acteurs ayant le plus fortement contribué au projet, acteurs que nous désignons par leurs initiales par souci de confidentialité, à l'exception de Florence Amardeilh, en charge du couplage entre IDE et ITM chez Mondeca, dont nous citons les travaux par ailleurs :

##### 3.4.1.1 Chez Temis :

Les acteurs impliqués dans le projet PressPro étaient les suivants :

- **Sylvie**, Temis Paris, linguiste et informaticienne, chef de projet pour la partie Temis, a consacré une grande partie de son temps au projet PressPro.
- **Amandine**, Temis Paris, linguiste et informaticienne, a travaillé comme développeuse sur le projet PressPro au développement informatique de la cartouche de connaissance, auquel elle a consacré une grande partie de son temps.
- **Fabienne**, Temis Paris, linguiste et informaticienne, a joué le même rôle qu'AS.
- **Christian**, Temis Paris, informaticien, a travaillé sur la partie catégorisation/indexation du logiciel, il a assumé le rôle de « chef de projet catégorisation »,

- **Jean-Pierre**, linguiste qui travaillait précédemment chez Xyleme, a assumé le rôle de « chef de projet cartouche de connaissance » et « expert linguiste », et, à ce titre, a défini les patrons d'extraction.

A côté de ces acteurs, d'autres ont joué un rôle plus ponctuel, ne consacrant qu'une faible partie de leur temps au projet :

- **Vincent**, Temis Grenoble, ingénieur avant-vente, informaticien de formation, qui a participé à la rédaction des spécifications, et ensuite transmis la responsabilité du projet à SG. Il a gardé un rôle dans le projet, au niveau de la relation commerciale avec PressPro et de la validation des développements effectués par Temis (travail appelé « qualité »).
- **Karene**, Temis Grenoble, linguiste et informaticienne, qui a travaillé sur les spécifications du projet.
- **Olivier**, Temis Grenoble, Core Product Technical Manager, responsable du suivi et de l'évolution des logiciels de Temis. Il a participé à la rédaction des spécifications.

#### 3.4.1.2 Chez Mondeca :

Les acteurs majeurs du projet PressPro chez Mondeca étaient :

- **Benoit**, informaticien, chef de projet depuis début 2003,
- **Florence Amardeilh**, informaticienne, qui a effectué sa thèse de doctorat en informatique sur le couplage entre moteur d'extraction et système de représentation des connaissances. C'est dans ce cadre qu'elle a travaillé sur les projets PressPro et Exinis. Elle a joué le rôle d'« intégrateur » du couplage IDE/ITM à partir de novembre 2003, selon la terminologie employée par Mondeca, i.e. de responsable de couplage ou chef de projet couplage.
- **Bernard**, chercheur, « expert ontologue », qui a joué le rôle de « chef de projet ontologie », et supervisé le travail sur le couplage effectué par F. Amardeilh.
- **Thomas**, informaticien, responsable du développement du produit ITM, ou chef de produit ITM, à partir de décembre 2003. Il a aidé à définir l'architecture de la passerelle OntoPop, pour que ce composant soit compatible avec l'architecture des autres composants d'ITM.

- **Anh**, informaticienne, a travaillé avec F. Amardeilh et Thomas au développement des composants logiciels de la passerelle.

A côté de ces acteurs, d'autres ont joué un rôle plus ponctuel, ne consacrant qu'une faible partie de leur temps au projet :

- **Gilles**, responsable commercial, qui a assuré la coordination avec le client et Vincent de Temis et Patrick de Xylème pour les phases de négociations entre les partenaires (répartition des budgets) et les phases contractuelles avec le client.
- **Jean Delahousse**, président, qui supervisait l'avancement du projet.

#### 3.4.1.3 Chez Xyleme :

Une personne était impliquée chez Xyleme pendant la majeure partie du projet, avant qu'une deuxième ne soit recrutée pour le développement de l'application documentaire :

- **Patrick**, chercheur en informatique, responsable technique de Xyleme, seul acteur de Xyleme initialement impliqué dans le projet PressPro. Il fut tout d'abord en charge de la récupération de la base d'articles et d'annotations existante, à la fois « chef de projet Xyleme » et « chef de projet logiciel » pour les trois membres du consortium. Ensuite, il aussi assumé la responsabilité du développement de l'application documentaire, alors que toute son activité n'était pas dédiée au projet PressPro.
- **Nicolas**, informaticien, a été recruté en fin de projet par Xyleme, pour développer l'application documentaire qui avait été spécifiée par PF.

#### 3.4.1.4. Chez PressPro :

Les personnes impliquées dans le projet chez PressPro, étaient :

- **Catherine**, la chef de documentation et chef de projet : elle a acquis une longue expérience de documentaliste dans divers groupe de presse, suivie d'une expérience de chef de documentation dans un grand journal économique français, avant d'intégrer PressPro.
- **Fabrice**, informaticien, responsable informatique du service documentation de PressPro, ou « responsable de l'informatique documentaire », responsable informatique du projet,

- **Yann** et **Philippe**, deux informaticiens initialement en charge du développement de l'application documentaire et de son intégration,
- les **documentalistes** elles-mêmes qui ont toutes participé à des groupes d'utilisateurs ou *user groups* à partir de mai 2004, et ce jusqu'à leur suspension brutale. Parmi elles, deux documentalistes dédiées ont ensuite travaillé dans les locaux de Temis, après l'arrêt de ces *users groups*, et ce essentiellement avec les chefs de projet de Mondeca et de Xyleme.

### 3.4.2 Un tissu d'interactions dense :

#### **Les collaborations internes des équipes-projet à partir de mai 2004 :**

Ces collaborations ont donné lieu à des réunions fréquentes chez Temis, réunissant les linguistes, travaillant sur le projet PressPro, qui étaient pratiquement tous dans les bureaux parisiens. Par ailleurs, il y a eu des réunions téléphoniques hebdomadaires d'environ une heure, réunissant tous les gens impliqués chez Temis, aussi bien à Paris qu'à Grenoble : ces « *conf call* » étaient l'occasion de faire le point sur les livrables, l'avancement des différentes tâches, le respect des délais et de la qualité. Chez Mondeca, les collaborations impliquaient des réunions internes toutes aussi fréquentes que chez Temis, la différence principale venant de ce que tous les acteurs du projet étaient sur le même site. Chez Xyleme, dans un premier temps, une seule personne a travaillé réellement sur le projet, tout en assumant théoriquement la tâche de responsable du projet pour les trois éditeurs. En fait, cette dernière tâche est restée assez théorique, la coordination entre 179 éditeurs à un niveau globale restant très faible. Par la suite, Xyleme a repris la tâche de développement de l'application documentaire (avec surtout le développement des interfaces utilisateur) et d'intégration de la solution. Pour cela, un informaticien a été recruté, qui a principalement travaillé avec le chef de projet Xyleme qui avait spécifié cette application.

#### **Les réunions avec les *user groups* de PressPro de mai à décembre 2004 :**

A partir de mai 2004, les éditeurs ont commencé ce qu'ils appelaient des « chantiers métiers », qui consistaient à travailler avec des groupes de documentalistes (*user groups*) pour comprendre leur métier et leurs besoins. En effet, les spécifications rédigées et remises à PressPro en mars 2004, n'étaient pas très détaillées, et le travail avec les documentalistes devait permettre de leur donner la finesse nécessaire. Il y avait plusieurs groupes de travail, sur des parties différentes de la solution à développer : un groupe travaillait sur le thesaurus, un autre sur les fonctionnalités de recherche souhaitées par les documentalistes, etc.... En tout il y avait huit groupes de travail,

regroupant cinq ou six documentalistes chacun. La plupart des cinquante documentalistes participaient donc à un de ces groupes de travail. Pour certains chantiers, les réunions étaient hebdomadaires, comme pour « ontologie » et « base de connaissance ».

Le but des réunions avec ces groupes était d'améliorer le développement des applications par le consortium, sur des aspects comme la création des « écrans métier » et l'enrichissement du thésaurus dans la perspective de l'indexation automatique. Tous les documentalistes participaient à l'un de ces groupes, et y jouaient théoriquement le même rôle. Mais il est apparu des leaders, qui étaient les documentalistes les plus motivées par le projet. En pratique, ces groupes d'utilisateurs au nombre de huit, présentaient un triple intérêt pour les chefs de projets des éditeurs du consortium :

1. Apprendre à connaître la manière de travailler des documentalistes, leurs besoins et leurs problèmes, pour définir une application documentaire, i.e. des interfaces utilisateur adaptées,
2. Apprendre à connaître la « presse people » pour développer une cartouche de connaissance et créer une ontologie personnalisée,
3. Faire de la pédagogie auprès des documentalistes, rôle que les chefs de projet du consortium appelaient « évangélisation ». L'importance d'expliquer les apports futurs de la solution en cours de développement, ne leur avait pas échappé, cela constituait même une de leurs priorités.

Les réunions des *user groups* impliquaient à chaque fois l'un des chefs de projet du consortium, qui participaient donc chacun à plusieurs de ces users groups, et parfois la chef de documentation. Ensuite, chaque chef de projet du consortium faisait un retour en interne, auprès des experts, informaticiens et autres acteurs impliqués dans le projet. Chaque chef de projet du consortium participait à deux ou trois groupes d'utilisateurs, et assistait donc à un assez grand nombre de réunions.

#### **Les réunions entre les trois éditeurs à partir de mai 2004 :**

Ces comités de pilotage réunissaient pendant environ deux heures, les chefs de projet des trois éditeurs chez PressPro, ils se déroulaient tous les quinze jours. Il s'agissait de faire le point sur l'avancement de chacune des briques logicielles, et de l'application documentaire en tant que telle (notamment des interfaces utilisateur).

### **La collaboration Mondeca/Temis à partir de mai 2004 :**

Cette collaboration était poussée, qui passait par des contacts et des réunions fréquentes, car Temis et Mondeca devaient coordonner leur travail avec précision. Temis et Mondeca travaillaient par itération : ils avançaient en parallèle, Temis travaillait à l'extraction d'information et définissait des règles ( patrons d'extraction), tandis que Mondeca définissait une ontologie, c'est-à-dire une modélisation de la connaissance contenue dans les articles de la presse people. Il fallait que progressivement les deux se rejoignent, car il y a une nécessaire correspondance entre les relations extraites par Temis et l'ontologie définie par Mondeca. Chacun travaillait de son côté, mais en tenant compte de l'évolution du travail de l'autre, en faisant converger progressivement leurs travaux respectifs. Par exemple, Temis reprenait une information selon laquelle « X s'est marié avec Y », et Mondeca vérifiait qu'une relation de type mariage était bien en cours de modélisation. Ainsi, parmi les nombreux échanges entre Temis et Mondeca, les plus poussés concernaient la chef de projet couplage et Amandine, chargée du développement de la cartouche de connaissance, à propos de la récupération des résultats fournis par ces cartouches. Ces échanges se déroulaient par téléphone, très régulièrement, plus d'une fois par semaine, et par mail. Ponctuellement, une réunion était organisée dans les locaux de Temis à Paris. Cela illustre le fait que la collaboration entre Temis et Mondeca a constitué un travail lourd, qui a permis au deux éditeurs de mieux se connaître.

### **La collaboration Xyleme/Temis :**

Cette collaboration a été nettement moins approfondie que la précédente. Temis estime avoir plutôt travaillé pour Xyleme, que collaboré avec Xyleme, puisque Temis a aidé Xyleme à reprendre l'existant (les articles du fond documentaire de PressPro). Temis en avait absolument besoin, et comme Xyleme avait du mal à avancer sur ce point, Temis l'a fait. Finalement Temis a dû consacrer du « temps homme » supplémentaire, sans contrepartie de la part de Xyleme.

### **La collaboration Xyleme/Mondeca :**

Le chef de projet Xyleme en charge du développement de l'application documentaire et de son intégration, a travaillé avec le chef de projet Mondeca, et avec la chef de projet couplage. Cette dernière a notamment travaillé avec lui à l'intégration du logiciel de Xyleme, XML Server, dans ITM, pour permettre le transfert, depuis ITM, des annotations sémantiques vers le logiciel XML

Server. Tandis que c'est le chef de projet Mondeca, aidé de la chef de projet couplage, qui a collaboré avec le chef de projet Xyleme, pour que ce dernier puisse définir des interfaces de recherche permettant de lancer des recherches d'articles stockés par Xyleme à partir de la base de connaissance ITM. La chef de projet couplage a aussi collaboré avec le chef de projet Xyleme à la définition de l'écran de validation des annotations.

### *3.4.3 Les étapes du projet :*

Nous présentons maintenant les étapes du projet, par brique logicielle, puis pour le système complet, une fois les différentes briques logicielles intégrées à celui-ci, i.e. à partir de fin 2004.

#### **3.4.3.1 Les briques logicielles de Temis :**

Pour le projet PressPro, Temis a vendu les logiciels IDE (« extractor ») et IDK (« categorizer »), briques logicielles qui doivent être personnalisées. Dans le cas de ce projet :

1. Le travail sur IDE fait par Temis, consistait à créer une « cartouche de connaissance » ou « cartouche métier » (logiciel permettant d'extraire de tout article de presse les informations susceptibles d'intéresser les journalistes du groupe PressPro). Ce travail est estimé par Temis au minimum à 50 jours/homme, quand le domaine est nouveau pour l'éditeur, ce qui était le cas ici. Dans le cas du projet PressPro, la cartouche était considérée comme très particulière par Temis, car elle n'avait rien à voir avec les cartouches déjà réalisées pour l'intelligence économique, ou la gestion des CV.
2. Le travail sur IDK fait par Temis consistait à entraîner, et à modifier si besoin est, un moteur statistique sur les catégories du plan de classement de PressPro (estimation initiale minimum de 10 jours/homme de travail). C'est un travail plus simple et plus marginal, car non directement couplé avec le travail de personnalisation de Mondeca. Nous n'y reviendrons donc pas par la suite, si ce n'est pour souligner la similarité des problèmes rencontrés par cette brique logicielle, avec celle de la cartouche de connaissance « people » dans le cadre de ce projet.

Temis conçoit des « cartouches métier », qui sont des briques logicielles adaptées à un domaine particulier (exemple : « intelligence économique dans l'industrie pétrolière ») qui viennent s'interfacer avec le moteur d'extraction IDE. Pour développer ces cartouches, il faut une

connaissance des métiers en question, et pour cela, dans le cas qui nous intéresse, rencontrer les documentalistes et les responsables de la documentation. Pour PressPro, il s'agissait d'intégrer tous les noms de personnes traquées par PressPro, susceptibles d'intéresser leurs journaux et magazines, et de créer toutes les relations pouvant exister entre ces personnes (ex : mariage). Concernant les différentes cartouches développées, Temis travaille généralement sur des domaines génériques comme les concepts de veille stratégique ou concurrentielle, qui peuvent être réutilisés. Par contre, pour cette cartouche « people », Temis n'a pu reprendre que quelques éléments de cartouches existantes (notamment en Intelligence Economique).

Une cartouche représente un ensemble de concepts. Il existe un plan de classement des biographies chez PressPro, et la cartouche doit être le reflet de ce plan de classement. Ensuite, ce qui sera extrait d'un article par exemple « X se marie avec Y », constitue une relation qui sera extraite et stockée dans la base de connaissance de Mondeca.

Après analyse du plan de classement de PressPro, et de certains documents, les linguistes de Temis ont défini un certain nombre de concepts. Par exemple, le concept « couple » : en cliquant sur « couple », dans le logiciel, on met en évidence 247 relations relatives au concept « couple ». Et on peut tout de suite accéder aux textes où l'on trouve ces relations. Pour PressPro, tous les concepts étaient déjà définis dans un plan de classement, et les documentalistes faisaient manuellement le travail d'analyse du texte en lien avec ces concepts.

« J'aime », « J'étais amoureuse », « il vit une passion » etc. Toutes ces formes verbales sont à placer dans la cartouche et ensuite le logiciel les sélectionne automatiquement quand elles figurent dans un article, quelles que soient leur conjugaison.

Il y a aussi dans ce plan de classement les concepts de « naissance », « santé », « nécrologie ». Le détail de l'extraction (exemple : la *date* de naissance d'une actrice), et ce que l'on va envoyer dans les bases de connaissance de Mondeca. Pour « nécrologie » : « [...] Adhémar Dion, le père de Céline s'est éteint le 30 novembre », on peut voir le détail de l'extraction, tous les éléments qu'on envoie à Mondeca : « nom de la chanteuse », « prénom de la personne décédée », « lien de parenté avec la chanteuse », « date du décès ».

### **Le développement d'une préversion de la cartouche métier d'août à novembre 2004 :**

Pendant l'été 2004, les linguistes de Temis ont commencé à développer la cartouche personnalisée « people ». Les gens qui y ont travaillé chez Temis étaient : Fabienne et Amandine. Ce fut donc surtout un travail de linguiste. Ce travail réalisé de septembre à novembre 2004, s'est traduit par l'élaboration d'une cartouche et la préversion de cette cartouche a été livrée à PressPro fin novembre 2004.

Temis a livré à PressPro des résultats d'extraction, effectuées à partir d'articles de presse, pour montrer ce que le logiciel de Temis arrivait à extraire comme information à partir de ces articles, aux *user groups*, à Catherine et à Fabrice. Ce n'était donc pas un logiciel utilisable.

### **Le travail sur la « qualité » à partir de décembre 2004 :**

Temis voulait savoir si les informations extraites correspondaient bien à celles qui devaient l'être. Finalement, seule la responsable de la documentation a fait un feed-back, du fait des problèmes rencontrés avec les groupes de travail. Celui-ci a d'abord été jugé excellent par Temis, mais Catherine a ensuite fait certaines remarques négatives. Mais pour Temis, il ne pouvait s'agir alors que d'un problème de « fine tuning », d'optimisation de la cartouche de connaissance (quelles sont les informations non extraites et qui auraient dû l'être (le « silence ») ? Quels sont les problèmes posés par ce qui est extrait ? Il est apparu, par exemple, un problème, nouveau pour Temis, lié à la résolution d'anaphore : « Johnny Hallyday est récemment parti en tournée [...], il s'est marié [...]. » Il fallait que le système comprenne que « il », c'est Johnny Hallyday). De nouveaux échanges entre Temis et PressPro ont eu lieu (réunions donnant lieu à des comptes-rendus écrits, en décembre 2004 et début 2005) à ce propos.

#### **3.4.3.2 L'application documentaire par PressPro, de mai à décembre 2004 :**

L'application appelée « application documentaire », est en fait le logiciel qui doit rassembler les différentes briques logicielles des trois éditeurs, et proposer aux utilisateurs les interfaces de validation et de recherche d'information, ainsi que de constitutions de dossiers aux utilisateurs finaux. Initialement, il était prévu que cette tâche soit effectuée par des informaticiens de PressPro : étaient concernés un développeur dédié au projet (spécialement recruté en CDD), un chef de projet/développeur, et toujours Fabrice, qui devait suivre l'avancement du projet. Après avoir été formés par les différents éditeurs, les développeurs de PressPro devaient donc

développer les interfaces, après que celles-ci eurent été spécifiées grâce au travail avec les *user groups*. Il y a aussi eu la participation des gens, cinq personnes, de l'infrastructure informatique de PressPro pour calibrer le système requis : le matériel et le système informatique.

### 3.4.3.3 Les briques logicielles de Mondeca :

#### **La définition de l'ontologie :**

L'ontologue expert a travaillé avec des groupes d'utilisateurs, et la chef de documentation de PressPro, a la définition d'une ontologie, permettant de représenter la connaissance contenu dans les articles de la presse « people ». Il a ainsi pu proposer une première version de l'ontologie qui a été intégrée à la version V0, livrée à PressPro en décembre 2004. Comme nous l'avons vu, la définition de l'ontologie et la définition de la cartouche de connaissance sont deux tâches distinctes, mais qui doivent être conduites en cohérence, ce qui a conduit l'ontologue expert de Mondeca et le linguiste expert de Temis, a des échanges réguliers.

#### **La création de la passerelle OntoPop**

La chef de projet couplage de Mondeca a d'abord travaillé avec l'équipe informatique interne de Mondeca, au développement des briques logicielles constitutives d'une passerelle logicielle susceptible de relier, dans un premier temps, les logiciels IDE et ITM. Pour effectuer ce développement, la chef de projet couplage a régulièrement consulté le linguiste expert de Temis, et l'ontologue expert de Mondeca, ainsi que le responsable produit ITM de Mondeca. Le concept de Règles d'Acquisition des Connaissances ou RAC, a ainsi été défini qui permet de définir un couplage personnalisé entre IDE, et à terme tout type de moteur d'extraction, et ITM, et à terme tout type de système de représentation des connaissances. Nous renvoyons ici à la partie suivante, qui traite en détail de la passerelle et de la méthodologie OntoPop.

#### **Le paramétrage des RAC :**

Le paramétrage des RAC s'effectue, une fois l'ontologie et la cartouche de connaissances définies avec précision. Or, comme nous le verrons, cette cartouche de connaissance n'a finalement pas pu être finalisée lors de ce projet, ce qui a eu pour conséquence, que les RAC n'ont pu l'être non plus.

#### 3.4.3.4. Le système intégré (à partir de décembre 2004) :

##### **Une version V0 du système livrée en décembre 2004 :**

Une version V0, regroupant une préversion de l'application documentaire et des briques logicielles de Xyleme et de Mondeca, mais pas celles de Temis, a été livrée à PressPro en décembre 2004. Dans cette version du système, les logiciels de Xyleme et de Mondeca étaient de plus encore indépendants. Xyleme avait repris pratiquement l'ensemble du fond documentaire de PressPro constitué de plus de deux millions d'articles dans sa base XML Server, et, l'ensemble des thésaurus de PressPro avait été transféré dans le module présenté par Mondeca. Cette V0 était un logiciel testable par PressPro, le but étant avant tout de montrer que le « consortium » était capable de reprendre l'existant, de proposer une solution qui engloberait ce que faisait l'ancienne. En résumé, les éditeurs ont voulu montrer à PressPro leur capacité à récupérer ses données et à l'intégrer dans leur système. Mais cette version V0 montrait peu de choses relatives aux fonctionnalités du futur système, et ne montrait pas les liens entre les différentes briques logicielles. C'était donc une sorte de prototype permettant à PressPro, de comprendre la direction prise par le consortium et de commencer à formuler des remarques à ce sujet. La partie Temis était absente, tandis qu'il restait encore du travail à Mondeca pour finaliser l'ontologie du domaine, et que l'application documentaire était en suspens.

Cette version a été présentée en décembre 2004 à l'ensemble des documentalistes chez PressPro, par les chefs de projet des trois éditeurs. Lors de cette présentation, certaines documentalistes ont contesté l'objet du projet, et ont réussi à susciter une réaction de rejet de la part de l'ensemble des documentalistes à son égard. Cela a abouti à l'arrêt immédiat des *user groups*.

##### **L'application documentaire confiée à Xyleme (02/2005) :**

Jusqu'en janvier 2005, ce travail, qui devait durer six mois (200 à 300 jours/homme) avançait très lentement. L'équipe-projet de PressPro avait du mal à appréhender la solution dans son ensemble, alors que de nombreuses tâches visant à la définition des différentes briques logicielles, étaient conduites en parallèle par Mondeca, Temis et Xyleme, au travers de relations généralement bilatérales. La tâche des informaticiens de PressPro n'était effectivement pas aisée, car elle consistait à intégrer différents modules en cours de développement et de personnalisation, modules dont ils ne maîtrisaient a priori pas la technologie. Le chef de projet de PressPro, a été relevé de sa tâche en janvier 2005, et il y a eu ensuite un mois de flottement.

Aussi en février 2005, suite aux échanges entre la chef de documentation de PressPro et les différents acteurs du projet du consortium, PressPro souhaitant absolument que cette application fonctionne en juin 2005, Xyleme, avec l'appui de Mondeca, a pris la direction de ce projet (maîtrise d'ouvrage déléguée). Xyleme a alors estimé la charge de travail à 120 jours/homme. Ce travail a finalement pu être réalisé grâce à l'embauche d'un informaticien par Xyleme.

- Version Beta de la version V1.0 du système en juin 2005 :

### **Livraison en juin 2005 :**

Les deux briques logicielles de Temis n'avaient pu être livrées en décembre 2004. Le travail sur la catégorisation des articles qui avait débuté en avril 2004, a abouti en avril 2005. Le travail sur la cartouche d'extraction, qui avait débuté en décembre 2004, a abouti en Mars 2005. Ces tâches effectuées, les différentes briques logicielles ont pu être assemblées à l'application documentaire, pour aboutir à une version Beta de la solution en juin 2005. A partir de mai 2005, le second projet destiné à Exinis, que nous retracerons dans la partie 7 de ce chapitre, a été lancé en parallèle du projet PressPro : ce projet a conduit à des développements nouveaux dans les modules logiciels constituant la passerelle logicielle OntoPop par Mondeca. La chef de projet couplage a alors dû s'assurer que de tels développements touchant au cœur de ces modules pouvaient bien être repris dans la solution développée pour PressPro.

### **Phase de test jusqu'en octobre 2005 :**

Les trois éditeurs (Temis, Mondeca et Xyleme) ont alors formé les membres d'un groupe de test chez PressPro (composé de la chef de documentation et de quelques documentalistes) pour qu'ils puissent tester la solution. Ils ont aussi formé deux formateurs, consultants indépendants, qui devaient eux-mêmes, par la suite, former l'ensemble des documentalistes de PressPro. Pendant l'été 2005, Catherine et le groupe de tests de PressPro ont donc fait des remarques sur la version Beta du système, qui ont entraîné des corrections. Ces remarques ont concerné principalement :

La qualité de la catégorisation et de l'annotation des articles, cette dernière fondée sur la cartouche de connaissance.

L'interface utilisateur : l'ergonomie du système s'est avérée insatisfaisante, ce qui a pu être amélioré grâce à des modifications relativement mineures.

### **Le problème de la qualité de l'extraction et de la catégorisation des articles :**

Les premiers dysfonctionnements qui avaient sauté aux yeux des documentalistes effectuant des tests, concernaient les briques logicielles de Xyleme et de Mondeca, car cela concernait des fonctionnalités de base et des performances facilement mesurables (temps d'accès à la base, etc...). Pour que les documentalistes puissent faire un test complet de la solution, il fallait tout d'abord que ces problèmes soient réglés. Donc, les briques logicielles de Temis sont restées « invisibles » aux yeux des documentalistes jusqu'au mois d'octobre 2005. En septembre 2005, il y a eu un grand nombre de corrections de bugs, finalement mineurs, par Mondeca et Xyleme, tandis que Temis n'était pas encore sollicité.

A partir de ce moment-là, les problèmes majeurs ont concerné les résultats d'extraction et de catégorisation des briques logicielles de Temis, le taux implicitement attendu de 80% d'indexation et de catégorisation pertinentes n'étant pas respecté.

Des tensions sont alors apparues entre la chef de documentation de PressPro et les acteurs du projet de Temis. La première reprochant les mauvaises performances des briques logicielles d'extraction et de catégorisation, les seconds reprochant à la chef de documentation de PressPro, son incompréhension du fonctionnement de ces briques. Une partie du contentieux est venue de ce que PressPro a testé les briques d'extraction et de documentation sur des articles de la presse traditionnelle, comme Le Monde, et non pas sur ceux de la presse « people » à partir desquels Temis avait travaillé. Pour Temis, il était normal que dans ces conditions ses briques logicielles fonctionnent imparfaitement, alors que pour PressPro cela n'était pas acceptable. Le responsable de l'informatique documentaire ayant quitté son poste, la chef de documentation était alors la seule responsable du projet chez PressPro, et à ce titre d'autant plus fragilisée par ce qui était perçu chez PressPro, comme l'échec des briques logicielles de Temis. Son travail dans le cadre du projet a alors été remis en cause par sa hiérarchie, et elle a quitté à son tour son poste en mars 2006. In fine, les chefs de projet du consortium ont estimé que la chef de documentation et le responsable informatique ne possédaient pas les compétences requises pour comprendre précisément et mener un tel projet.

#### **- Pré-production de novembre 2005 à février 2006 :**

La phase de pré-production suit la phase de test, et précède la phase de production en tant que telle, phase qui voit le système utilisé par toutes les documentalistes. Le passage de la phase de

pré-production à la phase de production se fait normalement après le « recettage » de la solution, i.e. la reconnaissance par le client que la solution développée correspond bien à ce qui avait été décrit dans le cahier des charges. La recette est définie par un cahier de recette qui définit le périmètre de la recette, les fonctionnalités à tester, la qualité à atteindre, etc. Normalement un tel cahier est rédigé au début du projet, mais dans le cas du projet PressPro, cela n'a pas été le cas.

Le système est rentré en phase de pré-production chez PressPro, en novembre 2005. Auparavant, le groupe de test de PressPro (Catherine et deux documentalistes) avait commencé à lister des bugs. Fin novembre 2005, cette recette n'était pas encore finalisée. Concernant la recette, le problème de Temis était différent de celui de Xyleme ou de Mondeca :

- Les briques logicielles développées par Xyleme et de Mondeca, qui proposent des fonctionnalités de transfert, validation, recherche, publication d'information, sont plus faciles à évaluer par le client : la fonction est-elle effective ou pas ?
- Par contre, dans le cas des briques logicielles de Temis, il s'agit de faire une évaluation qualitative de l'extraction et de l'indexation effectuée sur différents corpus documentaires. Il faut déterminer les taux d'extraction et d'indexation pertinents qui sont atteints, sachant que ces deux tâches sont effectuées habituellement par des opérateurs humains et aboutissent donc à des résultats relativement hétérogènes.

A partir de novembre 2005, chez PressPro un « comité de qualité » a été chargé de faire la liste des bugs. Il comprenait la chef de documentation, quatre documentalistes de PressPro réellement impliquées dans le projet, et le directeur technique de Xyleme. Globalement, les documentalistes ont jugé que la qualité du traitement des articles par le logiciel de Temis n'était pas satisfaisante, et qu'elles « ne voyaient pas à quoi servait Temis ! ». Par contre, les problèmes relatifs à l'application documentaire étaient quant à eux résolus et cette application donnait satisfaction à PressPro.

#### - Un système incomplet en production de février à juillet 2006 :

De février à juillet 2006, l'application documentaire a été mise en production, mais sans les briques logicielles de Temis. Celles-ci ne donnaient pas satisfaction, et Temis a dû les retravailler pendant cette période. Or, sans celles-ci, l'automatisation recherchée (d'annotation et de peuplement d'ontologie) devenait impossible, et seule subsistait une application permettant

d'annoter/peupler manuellement, puis de stocker les articles au format XML, pour ensuite faire des recherches et publier de dossiers.

#### **Installation de la « cartouche » Temis et recette en février 2006 :**

Le comité de qualité de PressPro, s'est réuni pour faire le point sur les bugs et sur les fonctionnalités, et a émis une recette « avec réserves », portant sur l'ensemble du projet, fin décembre 2005. Le service informatique de PressPro a ensuite installé une nouvelle version de la cartouche de connaissance de Temis en février 2006, et sans que celle-ci ait réellement été testée, la recette « définitive » de la solution a été validée par le « comité de qualité » de PressPro en février 2006.

#### **Fin du consortium et départ de la chef de documentation de PressPro, en mars 2006 :**

En mars 2006, le consortium a cessé d'exister, la recette de la solution ayant été signée par PressPro. A peu près au même moment, Xyleme a déposé son bilan. Une période de flottement s'en est suivie. Les informaticiens de PressPro ont repris le contrôle du projet, travaillant avec les chefs de projet de Mondeca et de Temis, à l'amélioration de la solution en production. Dès lors, ce travail n'a plus donné lieu qu'à des collaborations parallèles Temis/PressPro et Mondeca/PressPro. Au même moment, en mars 2006, la responsable de la documentation a quitté son poste, très certainement sanctionnée pour ce qui était perçu par la hiérarchie de PressPro comme l'échec du projet, quand bien même la solution était en production au service documentation. La question de l'efficacité des briques logicielles de Temis n'étant toujours pas résolue, la chef de projet de Temis a alors travaillé à la comparaison des résultats d'extraction et de catégorisation obtenus « manuellement » par les documentalistes avec ceux obtenus par les logiciels de Temis.

#### **- Le module Temis est débranché en juillet 2006 par PressPro :**

En juillet 2006, Temis a « débranché » les logiciels de Temis de l'application documentaire. Peu de temps auparavant, les documentalistes avaient écrit à la direction de PressPro, en affirmant l'inutilité des outils de Temis. En dehors de ce problème, le système était devenu stable et performant pour l'archivage et la recherche d'articles. Toutefois, le management de PressPro était dubitatif quant à la future maintenance de la solution, du fait de la disparition de Xyleme,

Mondeca ne s'estimant pas capable ni désireux d'assurer la maintenance de la brique logicielle de Xyleme.

**- L'arrêt complet du système fin 2006 :**

Le système fonctionnait avec les modules de Xyleme et de Mondeca, et les documentalistes avaient commencé à se l'approprier. Mais rapidement, PressPro a décidé de débrancher complètement la solution. La disparition de Xyleme, et les problèmes rencontrés avec Temis, ne laissaient à PressPro que Mondeca comme partenaire, susceptible d'assurer le suivi et la maintenance de la solution. Or Mondeca ne pouvait ni ne souhaitait assurer la maintenance de l'application documentaire. Finalement les responsables de la documentation de PressPro ont préféré tout arrêter, et choisi de transférer leurs articles et annotations dans une base de données classique.

### **3.5 Synthèse –Résumé du projet PressPro :**

**Une approche exploratoire, des apports multiples :**

Le projet PressPro a permis en tant que premier projet, de passer d'une approche théorique à une approche concrète de l'automatisation de l'annotation sémantique et du peuplement d'ontologie. Ainsi, le projet PressPro a été l'occasion de concevoir une plate-forme logicielle originale. Toutefois, cela ne signifie pas qu'il a permis une définition précise de cette plate-forme. Fondamentalement ce projet a permis d'en définir l'architecture, dont le cœur est formé des logiciels couplés d'extraction d'information et de représentation des connaissances. Cette plate-forme a été conçue pour pouvoir s'interfacer avec d'autres types de logiciels, comme celui de Xyleme, autorisant l'archivage de fichier au format XML. Elle a aussi été conçue de manière à permettre le choix d'un autre moteur d'extraction d'information qu'IDE de Temis, voire un autre système de représentation des connaissances qu'ITM de Mondeca. Le projet PressPro a permis le développement de la passerelle OntoPop, permettant le couplage entre IDE et ITM au travers de la conception de RAC (sur lesquelles sur nous reviendrons dans la partie 4 de ce chapitre) et du langage Opal permettant leur rédaction. Il a aussi permis aux acteurs impliqués de ressentir le besoin d'une méthodologie relative à la définition du couplage évoqué (cf. points 4, 5 et 6 de ce chapitre, relatifs à la passerelle et à la méthodologie OntoPop).

Le projet qui présentait un côté exploratoire, a donné lieu à des tâtonnements, et à de nombreuses itérations. Les éditeurs Temis et Mondeca estiment que le travail effectué dans le cadre du projet PressPro était proche d'un travail de recherche : « il n'existait rien, tout était à concevoir et à développer » (entretien avec F. Amardeilh, le 20/02/2008). Ainsi, il n'y a pas eu de prototype de la solution qui aurait permis de faire réagir les utilisateurs, alors que c'est l'usage dans l'industrie du logiciel, car l'architecture de la solution n'était alors pas définie.

### **Une architecture logicielle nouvelle :**

Le projet PressPro, a permis de définir les bases de l'architecture logicielle, les briques logicielles qui devaient la constituer. Cette architecture a ensuite été redéfinie, et précisée pour le projet Exinis. La solution développée pour PressPro, a en effet bénéficié des apports du projet Exinis, car ces deux projets ont été partiellement concomitants. Le projet PressPro a permis de définir l'architecture globale du processus d'annotation et de peuplement, les différentes composants logiciels, les flux de données en entrée et en sortie de chaque module, et la connexion sous forme de plug-in externe de différents logiciels tels que les systèmes de gestion de contenu (CMS) comme celui de Xyleme, ou les moteurs de recherche.

### **Un premier projet qui pose la question d'une double-intégration :**

Le projet PressPro fournit un premier champ d'application au couplage des technologies d'extraction terminologique et de représentation des connaissances. Il pose la question majeure de ce que nous appelons la double-intégration de la solution : l'intégration de la solution globale et du couplage IDE/ITM.

Le premier aspect est lié au problème de l'absence d'un acteur chargé de l'intégration globale de la solution, considéré comme la « principale erreur de ce projet » par les chefs de projet de Temis et de Mondeca : les éditeurs du consortium en réclamaient un, PressPro ne voulait pas et pensait pouvoir réaliser cette tâche seul. Mais finalement, PressPro n'a pas été capable d'accomplir cette tâche, et celle-ci a été attribuée aux membres du consortium. Ceux-ci ne voulaient a priori pas le faire car cela n'était tout simplement pas leur métier. Hélas, la solution était trop innovante pour qu'un intégrateur, une SSII, s'engage sans risque à en réussir l'intégration. Il apparaît qu'une des difficultés propres au déploiement d'innovations logicielles radicales et modulaires, réside dans la difficulté à trouver un acteur susceptible d'en réaliser l'intégration informatique. Cela résulte du

manque de compétences des intégrateurs, en général, sur les technologies innovantes employées et sur le couplage de ces technologies, et de leurs réticences à développer de telles compétences sur des niches de marché de taille forcément restreinte au départ.

Le deuxième aspect est lié quant à lui au couplage entre logiciel d'extraction de l'information et logiciel de gestion des connaissances. Cette tâche nécessite un « chef de projet couplage », tel qu'il sera défini ultérieurement dans la méthodologie OntoPop, qui permet d'organiser les multiples interactions et de définir les livrables relatifs à cette tâche.

### **Le succès du couplage entre modules d'extraction d'information et de représentation des connaissances :**

L'approche OntoPop, notamment le système de Règles d'Acquisition de Connaissance mis en place dans ce projet, s'avère performante même si la pertinence des extractions du moteur d'extraction d'information n'était pas satisfaisante pour PressPro. Le projet PressPro a ainsi permis à Mondeca, surtout, et à Temis de développer les compétences, liées au couplage de leurs logiciels respectifs, et de créer une offre commune de « création automatisée de base de connaissance » destinée au secteur de la presse et de l'édition. La réussite du couplage entre IDE et ITM, grâce à la passerelle logicielle OntoPop, a ainsi abouti à la conclusion d'un partenariat commercial entre Temis et Mondeca.

### **Des relations multiples, des interactions foisonnantes entre les acteurs du projet :**

Un foisonnement d'interactions entre acteurs de différentes organisations n'ayant pas collaboré auparavant, a permis l'émergence d'une solution logicielle innovante. Elles résultent de l'instauration d'un certain nombre de relations interorganisationnelles, nécessaires au déroulement du projet :

1. Au départ, des relations se sont instaurées entre Catherine, la chef de documentation et les responsables commerciaux de Temis et Mondeca.
2. Il y a ensuite eu des relations entre les ingénieurs avant-vente (jouant en quelque sorte le rôle de chef de projet) des membres du consortium, d'une part, Catherine et Fabrice, le responsable de l'informatique documentaire de PressPro, d'autre part, et ce à partir de la phase de rédaction des spécifications.

3. Après cela, des relations se sont instaurées entre d'une part les chefs de projet de chaque membre du consortium et, d'autre part, les *user groups* de PressPro, Catherine et Fabrice, pour apprendre le métier des documentalistes et leur façon de travailler, afin de pouvoir personnaliser les différents logiciels de la solution.
4. Il y a eu en parallèle, des relations entre les membres d'équipe-projet des membres du consortium et les informaticiens de PressPro en charge du développement de l'application documentaire.
5. En parallèle aussi se sont développées des relations inter-éditeurs, visant à relier les différents logiciels proposés par chacun d'entre eux (IDE/ITM, ITM/logiciel de Xyleme, IDE/logiciel de Xyleme), et ainsi à concevoir la passerelle de couplage entre logiciel d'extraction d'information, et système de représentation des connaissances, qui constitue un projet à l'intérieur du projet,
6. Il y a ensuite eu dans la phase de test de la version Beta, puis de la version 1.0, des relations entre les chefs de projet des membres du consortium et le groupe de test de PressPro.

Ces différentes relations ont généré un grand nombre d'interactions interorganisationnelles, qui présentent des caractéristiques diverses, quant aux acteurs concernés, et à l'atmosphère de ces interactions (Hakansson, 1982), notamment. Ce foisonnement d'interactions était difficile à organiser, ce qui a été souvent mal ressenti par les acteurs du projet, qui ont eu l'impression de travailler inefficacement et de perdre beaucoup de temps. Les acteurs du projet chez Temis et Mondeca ont ainsi eu l'impression qu'il manquait un véritable chef de projet. De ces insatisfactions naîtra la méthodologie OntoPop.

### **La faiblesse des liens entre membres de l'équipe-projet : une organisation faiblement couplée.**

Le fonctionnement du projet s'est avéré extrêmement souple, dès qu'un blocage ou un problème de nature à en perturber le cours est survenu, une solution a été trouvée dans des délais brefs, indiquant une réelle motivation de la part des éditeurs de logiciel appartenant au consortium ainsi que du client PressPro. A la fin, hormis le fait que les briques de Temis ont été débranchées, le projet n'a eu *que* six mois de retard : lorsqu'il a fallu réattribuer le lot « développement de l'application documentaire » suite aux difficultés de PressPro, lorsque Xyleme a eu des problèmes

pour récupérer les données antérieures, lorsque les documentalistes ont refusé de poursuivre leur participation aux *user groups*, etc.... Tout cela s'est déroulé dans un contexte où les organisations, parties prenantes du projet, étaient faiblement couplées. Le « consortium » mis en place par les trois éditeurs, ne correspondait en fait à aucun partenariat formel, et s'est dissout dès le recettage du projet effectué par PressPro. Il semble que le projet n'ait pas été réellement dirigé, Catherine n'a pas pu assumer cette tâche, et personne ne semble l'avoir assumée à sa place. Différents pôles de coordination sont apparus, autour du couplage IDE/ITM, autour du développement de l'application documentaire, mais pas au niveau global.

### **La relation de chaque éditeur du consortium avec les documentalistes :**

Nous pensons qu'il est utile de revenir sur les interactions entre les éditeurs du consortium et les documentalistes, futures utilisatrices du système. A la différence de Xyleme et Mondeca, Temis est le seul éditeur à ne pas avoir pu travailler directement avec celles-ci. Les raisons en sont multiples : tout d'abord, il y a le fait que les logiciels de Temis sont en « back office », que le premier niveau d'utilisation du système complet concerne uniquement Xyleme et Mondeca : pour « voir » la solution Temis, dans le cadre du système global, il fallait que les briques logicielles de Xyleme, de Mondeca et surtout l'application documentaire soit en place : ainsi les futures utilisatrices étaient dans l'impossibilité de simuler l'utilisation des briques développées par Temis, à l'inverse de celle développées par Xyleme ou Mondeca. Toutefois, Temis avait la possibilité de faire évaluer des résultats d'extraction. On peut aussi penser que le bouleversement des habitudes de travail qu'allait entraîner la mise en œuvre de l'indexation/catégorisation/extraction par les logiciels de Temis, a fait peur aux responsables de projet chez PressPro : cette inquiétude a largement été partagée par l'équipe-projet de Temis, qui craignait que les documentalistes se mettent en grève dès la phase de test. Tout cela a conduit à ce que seule, Catherine, chez PressPro, travaille directement avec Temis sur ce projet. Ainsi, pendant la plus grande partie du projet, l'équipe-projet de Temis a eu le sentiment que tout se passait bien de son côté, que sa solution allait être opérationnelle, et que seuls des problèmes sociaux pouvaient les inquiéter. Mais, finalement, ce n'est que vers la fin du projet, alors que des premiers tests concrets étaient effectués en interne chez PressPro, que les limites de la solution sont apparues, alors que les problèmes liés aux logiciels de Xyleme et Mondeca se résolvaient progressivement. On peut considérer que le manque de contacts directs entre l'équipe-projet de Temis et les futures utilisatrices a énormément nuit à la remontée des problèmes inhérents à l'extraction terminologique.

### **Le rôle du client et l'évolution du leadership du projet : deux phases successives.**

1. **Le client architecte du projet tente de le diriger :** lors d'une première phase (de juin 2003 à décembre 2004), c'est PressPro, par l'intermédiaire principalement de la chef de documentation, qui a dirigé le projet : Elle en a eu l'idée, elle a contacté Mondeca et Temis, les a conduits à s'associer et ensuite à intégrer Xyleme, malgré leurs réticences. De plus, les informaticiens de PressPro ont reçu la responsabilité du développement de l'application documentaire, qui devait agréger les différentes briques logicielles. Enfin, les échanges entre les documentalistes réunies au sein des *user groups*, et les chefs de projet de chaque éditeur de logiciel du consortium se sont déroulés chez PressPro, sous la supervision la chef de documentation.
2. **Les éditeurs de logiciel conduisent le projet :** dans une deuxième phase, (février 2005 à février 2006), le leadership du projet a été assumé par les trois éditeurs du consortium, et a semblé revenir plutôt au binôme Mondeca/Xyleme, qui a repris le projet de développement de l'application documentaire. Les informaticiens de PressPro en charge du développement de l'application documentaire, ne connaissaient pas les différentes technologies proposées par Temis, Mondeca et Xyleme. Cette lacune semble les avoir empêchés de véritablement « se saisir » du projet. La tâche de gestion de projet, en tant que telle, n'a par ailleurs pas été clairement attribuée : d'un côté, PressPro, était chargé de l'application documentaire qui fédère les différentes briques logicielles, et de l'autre, Xyleme était en théorie le responsable du projet pour l'ensemble des trois éditeurs, en quelque sorte le chef de projet « consortium ». Cette répartition des rôles a entraîné certain flou néfaste au projet dans le sens où il manquait un véritable chef de projet, doté d'une vision globale. Elle a par contre permis une grande autonomie à chacun des acteurs, et donc laissé libre cours à tout un ensemble d'interactions, ayant in fine permis de bâtir la solution recherchée.

Nous relierons la question de l'évolution du rôle du client à celle de la compétence de la chef de documentation de PressPro, seule interlocutrice directe de Temis pendant toute une partie du projet. Il est étonnant de constater à quel point, cette compétence a été évaluée de manière fluctuante par les équipes-projet de Temis et de Mondeca. Au départ et pendant les trois quarts du projet, elle a été considérée comme une interlocutrice très compétente, avec laquelle il était facile de travailler, et en laquelle les membres du consortium avaient toute confiance. Il est vrai que c'est elle qui avait imaginé l'architecture de la solution. Elle bénéficiait aussi d'un crédit

certain, en tant qu'ancienne responsable de la documentation d'un grand journal économique. Toutefois, dans la dernière partie du projet, les membres du consortium ont perçu chez elle un manque de compétences et de compréhension du projet. A la réflexion, cela apparaît assez logique celle-ci n'ayant aucune expérience ni formation en gestion des connaissances, ni en *text mining*.

Nous allons maintenant décrire les principes sur lesquels reposent le composant logiciel destiné au couplage entre moteur d'extraction et système de représentation des connaissances. Ces principes, dont la notion de Règle d'Acquisition de Connaissances, ont été conçus pendant les projets PressPro, principalement, et Exinis, deuxième projet sur lequel nous revenons dans la partie 7 de ce chapitre. Ensuite, nous aborderons le fonctionnement de cette passerelle logicielle, appelée OntoPop, et décrirons les différentes briques logicielles qui la composent. Enfin, nous décrirons la méthodologie, elle-aussi appelée OntoPop, qui vise à décrire les phases successives de mises en œuvre d'une solution de « création automatisée de base de connaissance » comme nous avons choisi de la nommer, ou selon la terminologie employée par Mondeca d' « annotation sémantique et de peuplement d'ontologie ». Cette méthodologie a été appliquée au cours du projet Exinis, que nous décrirons ensuite. La conception de la passerelle OntoPop, de la notion de RAC, et de la méthodologie OntoPop, est principalement le fruit d'un travail effectué par F. Amardeilh, en collaboration avec d'autres acteurs de Mondeca, mais aussi de Temis et d'Exinis. F. Amardeilh a travaillé sur le sujet dans le cadre de sa thèse en informatique soutenue en 2007, et ce dès le début du projet PressPro. Elle a ainsi assumé au cours de projets PressPro et Exinis, le rôle de responsable de couplage. C'est pourquoi les éléments relatifs à la passerelle logicielle et à la méthodologie OntoPop que nous présentons maintenant, proviennent en grande partie du document de thèse de F. Amardeilh (2007), complété des éléments que nous avons collectés directement auprès des principaux acteurs concernés, soit F. Amardeilh elle-même, le chef de projet et le chef de produit ITM de Mondeca, l'architecte innovation d'Exinis, ainsi que le responsable technique *core product* et l'ingénieur avant-vente de Temis.

#### **4. LA PASSERELLE LOGICIELLE ONTOPOP ET LA NOTION DE RAC (REGLES D'ACQUISITION DES CONNAISSANCES) :**

Nous traitons ici de sujets très techniques, mais qui nous ont semblé indispensables à la compréhension du processus de modularisation que nous étudions. Comme nous l'avons dit, les

résultats et descriptions techniques que nous présentons maintenant sont, en grande partie, issus de la thèse en informatique de F. Amardeilh (2007).

#### 4.1 Introduction :

La démarche OntoPop constitue une contribution à la conception et à la réalisation d'applications pour l'annotation sémantique et le peuplement d'ontologie à partir de documents textuels. Elle s'appuie sur les théories, les méthodes et les techniques développées tant en Informatique Linguistique que dans le domaine plus récent du Web Sémantique. Les moteurs d'extraction d'information ne se sont presque pas intéressés à la manière de combler le fossé existant entre d'une part les représentations du contenu du texte par les outils d'extractions d'information, à savoir les arbres conceptuels, et d'autre part la représentation de ce même contenu par l'utilisation d'une ontologie.

Comme nous l'avons vu, le travail amorcé pour un laboratoire pharmaceutique, avait conduit Mondeca et Temis à faire un certain nombre de constats relatifs au couplage entre moteur d'extraction d'information et outil de représentation de connaissance. Cette réflexion a trouvé son prolongement dans le travail effectué pour le projet PressPro, principalement par la « responsable de couplage », et débouché sur la conception de ce que Mondeca appelle les « règles d'acquisition de connaissance » ou « RAC ». Ces règles se fondent sur les travaux de recherche relatifs à l'exploration contextuelle (Desclés et al. 1991, 1993), et plus précisément sur la notion de « règle d'exploration contextuelle » (Crispino, 2003). Nous allons tenter de décrire maintenant ce que sont ces RAC et comment elles ont émergé.

Fondamentalement, le couplage entre les logiciels IDE et ITM pose le problème de la correspondance entre différentes représentations d'un contenu textuel. En effet, les arbres conceptuels (issus des moteurs d'extraction d'information) constituent une représentation peu «sémantisée» du contenu textuel. Malgré la conceptualisation de certaines formes, ces arbres ne résultent que d'une analyse superficielle du texte qui n'appréhende pas la signification globale du contenu, mais en donne une vue linéaire, fragmentée en fonction de certaines unités textuelles. L'enjeu de l'innovation en ce qui concerne le couplage IDE/ITM est donc de mettre en place un mode d'interprétation de ces arbres permettant de proposer une représentation formelle du contenu, basée sur une ontologie de domaine. Mondeca en collaboration avec Temis a donc cherché à développer une passerelle entre ces deux modes de représentation, passerelle qui puisse

aussi être indépendante des outils utilisés, que ce soit les moteurs d'extraction ou l'outil de représentation des connaissances. La passerelle prend la forme d'un format intermédiaire de représentation couplé aux technologies issues du Web Sémantique. Ce format permet de définir un ensemble de règles, appelées par Mondeca, Règles d'Acquisition de Connaissance (RAC).

## 4.2 Une passerelle pour l'annotation sémantique et le peuplement d'ontologie :

Certains outils d'annotation sémantique proposent l'automatisation du traitement grâce à l'intégration d'un outil d'extraction d'information. Or, l'objectif poursuivi par Mondeca en accord avec Temis était de développer un outil d'annotation sémantique, ou de peuplement d'ontologie, indépendant du système d'extraction d'information utilisé, afin de proposer aux futures applications davantage de flexibilité. Par exemple, dans un cas nécessitant l'exploitation des coréférences, l'outil d'annotation pourrait être couplé avec GATE, logiciel open source, alors que pour l'exploitation de scénarios complexes dans un domaine d'application précis, l'outil d'annotation devrait être couplé avec IDE.

Mais pour cela, il faut trouver une solution générique au fossé existant entre les arbres conceptuels et la représentation formelle du contenu comme modélisée dans les ontologies (cf. Figure 9).

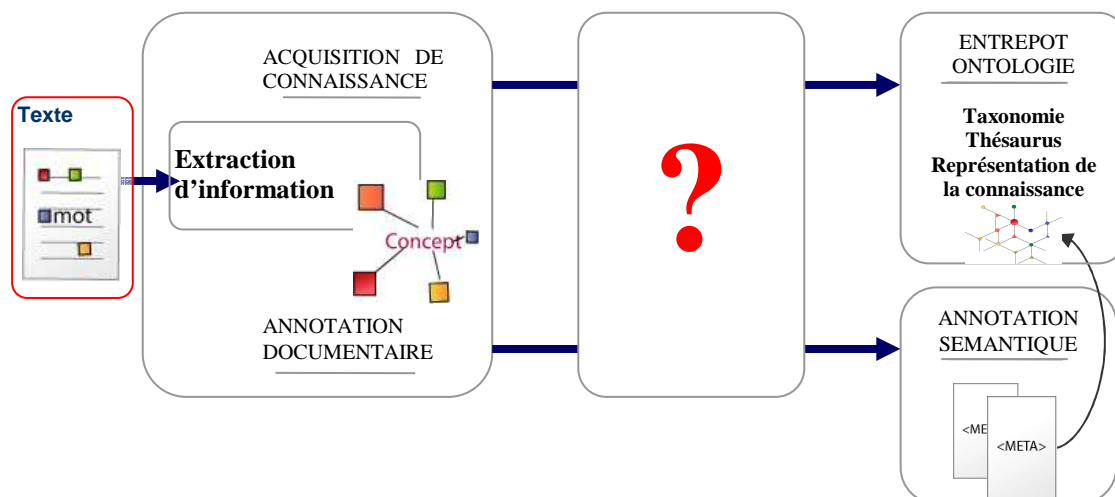


Figure 9 : Fossé entre la représentation textuelle et la représentation sémantique (Amardeilh, 2007).

Par conséquent, il est nécessaire de concevoir un niveau intermédiaire entre ces deux représentations qui permette le couplage des outils d'EI avec les outils d'annotation et/ou de peuplement d'ontologie.

Une des spécificités de la passerelle logicielle OntoPop consiste à réaliser cette correspondance grâce à la déclaration d'un ensemble de Règles d'Acquisition de Connaissance (RAC). Elles s'appuient sur des étiquettes sémantiques produites par les arbres conceptuels des outils d'extraction d'information et sur des éléments (concepts, attributs et relations) modélisés dans l'ontologie. Elles décrivent la manière dont un concept de l'ontologie va être instancié ou utilisé pour l'annotation documentaire. Pour cela, elles identifient l'étiquette sémantique qui déclenchera le processus d'instanciation ou d'annotation sur ce concept (figure 10).

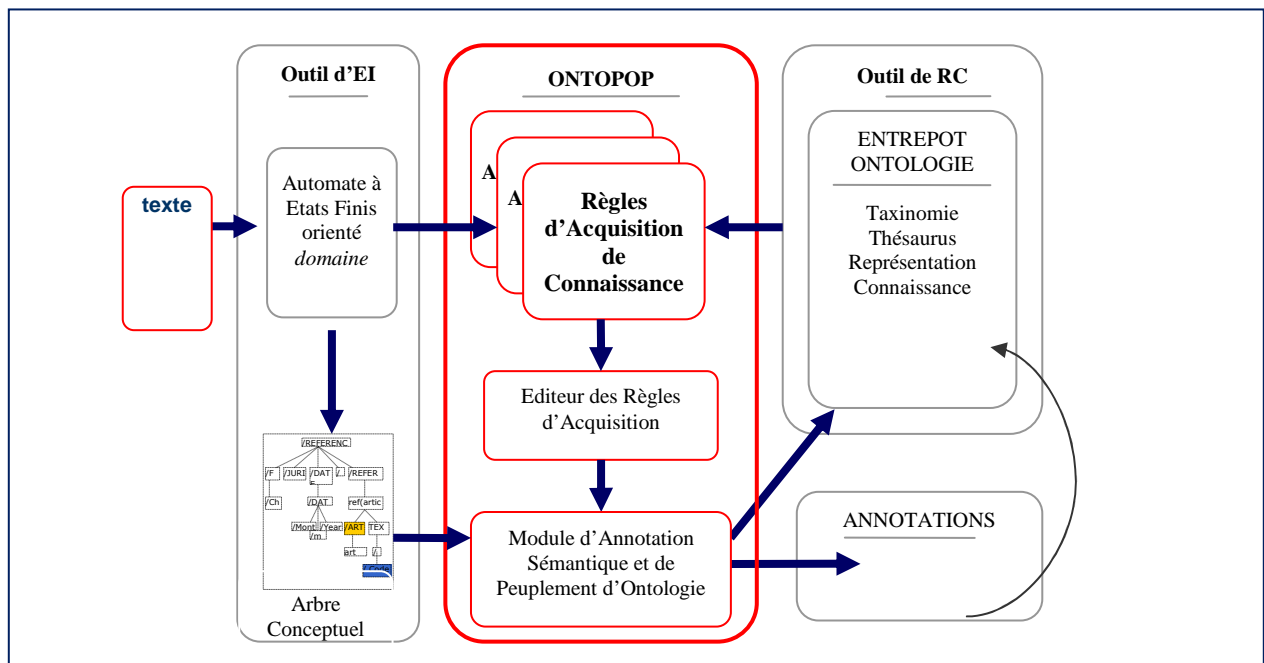


Figure 10 : La passerelle logicielle OntoPop (Amardeilh, 2007).

La démarche OntoPop propose également une méthodologie pour la mise en place d'une application d'annotation sémantique et/ou de peuplement d'ontologie basée sur la définition des RAC et une passerelle logicielle. Cette passerelle se compose d'un Editeur des RAC et d'un Module d'Annotation Sémantique et de Peuplement d'Ontologie. L'Editeur compile les RAC pour les traduire dans un langage machine afin qu'elles puissent être traitées par l'application documentaire. Le Module procède à la transformation des arbres conceptuels (cf. figure 7, p 149) en représentations formelles, que ce soit les annotations sémantiques de chaque document analysé ou les nouvelles instances à insérer dans la base de connaissance.

### 4.3 La formalisation des Règles d'Acquisition de Connaissance

#### 4.3.1 L'importance du contexte dans les arbres conceptuels :

Pour déterminer le format intermédiaire permettant de passer d'une représentation du texte sous la forme d'un arbre conceptuel à une représentation formelle basée sur la modélisation d'une ontologie, il faut comparer l'ensemble des étiquettes sémantiques constituant les arbres conceptuels d'un corpus et l'ensemble des éléments (concepts, relations, attributs) de l'ontologie du domaine. Dans l'exemple fourni par le projet PressPro, le domaine de la presse « People », parmi tous les éléments de l'ontologie, seules dix classes (4 correspondant aux entités nommées et 6 correspondant aux événements), 14 relations et 13 attributs doivent être couplés aux étiquettes sémantiques. Par exemple, la classe « Personnalité », sous-classe de « Personne », possède les attributs « date de naissance », « alias », etc. La classe « Article » possède les attributs « date de publication », « source », « auteur »... La classe « Article » est aussi reliée à la classe « Personnalité » à travers la relation « indexation personnalité », qui sera utilisée comme annotation sur le document.

Pendant la tâche de comparaison, toutes les informations nécessaires au couplage sont centralisées, comme dans le Tableau 1, page suivante. Les différents cas de couplage à traiter sont les suivants :

- Une seule étiquette sémantique correspond à un seul élément, cf. la classe « /Film ».
- Plusieurs étiquettes sémantiques correspondent au même élément, cf. la classe « Personnalité ».
- Une étiquette sémantique correspond à plusieurs éléments du même type, cf. l'étiquette « /COUPLE ».
- Une étiquette sémantique correspond à plusieurs éléments, de types différents, cf. l'étiquette « /ActorNamed ».
- Une étiquette sémantique ne correspond à aucun élément, cf. l'étiquette « Evenement\_Imminent ».
- Un élément ne correspond à aucune étiquette sémantique, cf. l'attribut « Signe Zodiacal ».

Élément dans l'ontologie	Type dans l'ontologie	Étiquette Sémantique	Contexte des étiquettes dans l'arbre conceptuel
Film	Classe	/OeuvreFilm	
Personnalité	Classe	/Personnalité	
		/NomPotentielDePersonne	
Mariage	Classe	/COUPLE	∃ Enfant = /Mariage
Divorce	Classe		∃ Enfant = /Divorce
Personnalité	Classe	/ActorNamed	∃ Enfant = /Personnalité
Conjoint	Relation		∃ Enfant = /Personnalité <b>and</b> ∃ Parent = /COUPLE
Indexation Personnalité	Relation		
Lieu de naissance	Attribut	/Location	∃ Parent = /DATE-NAISSANCE
Lieu du mariage	Attribut		∃ Parent = /COUPLE <b>and</b> ∃ Frère = /Mariage
Indexation lieu	Attribut		
Date de naissance	Attribut	/DATE	∃ Oncle = /Personne <b>and</b> ∃ Père = /Naissance <b>and</b> ∃ Ancêtre = /DATE-NAISSANCE
		/Evenement_Imminent	
Signe Zodiacal	Attribut		

**Tableau 1 : Tableau de comparaison éléments ontologiques versus étiquettes sémantiques (Amardeilh, 2007).**

Si une étiquette sémantique correspond à plusieurs éléments de l'ontologie, alors la prise en compte du contexte de cette étiquette dans l'arbre conceptuel est nécessaire pour résoudre les ambiguïtés. Le contexte d'une étiquette sémantique comprend ses ancêtres (parents, grands-parents et autres aïeux), ses descendants (enfants, petits-enfants, etc.), ses frères, ses oncles, etc. En fait, le contexte est l'ensemble des nœuds composant le sous-arbre auquel appartient l'étiquette sémantique étudiée (cf. figure 11, page suivante). La présence (« $\exists$ ») ou l'absence (« $\text{not}()$ ») d'un ou de plusieurs nœud(s) dans ce contexte permet de déterminer l'élément de l'ontologie qui peut être couplé à l'étiquette sémantique.

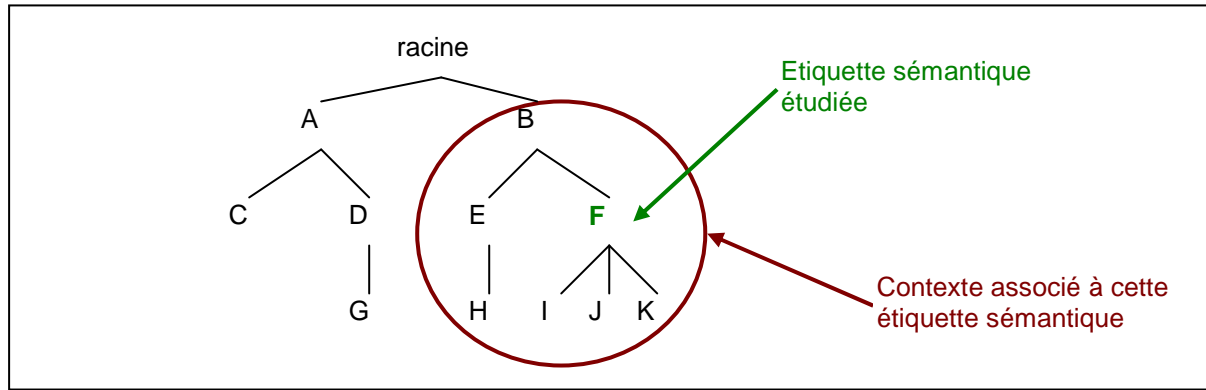


Figure 11 : Contexte d'une étiquette sémantique dans un arbre contextuel (Amardeilh, 2007).

Par exemple, si l'étiquette sémantique “/COUPLE” possède un nœud enfant “/Divorce”, un événement de la classe “Divorce” sera instancié. Par contre, si ce nœud enfant est le nœud «Mariage», alors il s'agira d'un événement de la classe « Mariage ». La notion de contexte est donc indispensable car le couplage entre les étiquettes sémantiques et les éléments de l'ontologie n'est que rarement une simple bijection. Ainsi, ce contexte doit être facilement appréhendable par les Règles d'Acquisition de Connaissance.

#### 4.3.2 Méthode et règles d'exploration contextuelle fondement des RAC :

Afin de construire des RAC capables d'appréhender le contexte des étiquettes sémantiques, Amardeilh (2007) s'est inspirée des travaux menés sur l'**exploration contextuelle** par Desclés (1991, 1993). La méthode repose sur l'identification d'indicateurs linguistiques qui explicitent une valeur sémantique à repérer dans une application donnée. Or, le repérage de ces indicateurs linguistiques n'est pas toujours suffisant pour l'identification de la valeur sémantique car comme le précise Minel (2003), « le rapport entre signifiants et signifiés n'est pas bijectif dans les langues ». Ainsi, outre l'identification des **indicateurs linguistiques**, il est nécessaire de repérer d'autres indices linguistiques, i.e. les **indices complémentaires**. Ces derniers conditionnent alors l'affectation de la valeur sémantique à l'unité textuelle analysée. Ces indices complémentaires peuvent être de nature très diverse comme une ponctuation, une position donnée dans l'unité textuelle analysée, un élément structurel du document comme un titre ou encore un type d'acte discursif comme une conclusion (Desclés, 1993).

La **méthode d'exploration contextuelle** fait donc intervenir, pour l'attribution d'une valeur sémantique donnée, un **indice déclencheur** et des **indices complémentaires** qui analysent le contexte de l'indice déclencheur pour déterminer l'attribution d'une valeur sémantique donnée.

L'objectif de ces « règles d'exploration contextuelle » consiste à formaliser la démarche proposée dans la méthode sous la forme : « si l'indicateur qui est identifié dans un texte T et si l'on constate la présence des indices Ip dans les contextes Cp alors prendre la décision Dj' » (Crispino 2003).

Par conséquent, **une règle d'exploration contextuelle** se divise en trois parties : 1) la partie « **déclenchement** » qui contient l'indice déclencheur et l'étiquette sémantique associée, 2) la partie « **conditions** » qui décrit les indices complémentaires à repérer ou non dans le contexte de l'indice déclencheur et 3) la partie « **actions** » qui n'est exécutée que si toutes les conditions sont vérifiées.

### *4.3.3 Des règles d'exploration contextuelle aux règles d'acquisition de connaissance :*

A l'origine du concept de **règle d'acquisition de connaissance**, se trouve l'application des principes de la méthode d'exploration contextuelle pour repérer les étiquettes sémantiques d'un arbre conceptuel, lui-même issu d'un moteur d'extraction. Pour repérer ces étiquettes, il faut explorer des ressources étiquetées, les arbres conceptuels, afin d'identifier des marqueurs, non pas linguistiques mais sémantiques, pour les rapprocher des éléments modélisés dans l'ontologie du domaine concerné. La notion de contexte dans ces arbres est également importante et décisive puisque c'est la présence ou non de certaines étiquettes sémantiques dans l'entourage du nœud déclencheur qui va permettre l'activation de la règle. Ce déclenchement provoque la création d'annotations sémantiques ou de nouvelles instances dans la base de connaissance. Mondeca a donc adapté les notions de REC (Règles d'Exploration Contextuelle) pour définir des Règles d'Acquisition de Connaissance (RAC).

### *4.3.4 Définition des constituants d'une Règle d'Acquisition de Connaissance : PressPro*

#### **4.3.4.1 Introduction :**

Les Règles d'Acquisition de Connaissance fournissent une solution adaptable visant à combler le fossé qui sépare les deux niveaux de représentation du contenu que nous étudions. Elles ont pu être définies principalement grâce au travail sur le projet PressPro, mais aussi sur le projet Exinis (cf. partie 7 de ce chapitre), qui a mis en évidence la nécessité d'enrichir le concept de REC, en introduisant une dimension supplémentaire, la partie « option ».

Mondeca définit une **Règle d'Acquisition de Connaissance** (figure 12) comme étant composée de **la classe du nœud indicateur N** (partie « Déclenchement »), d'un **ensemble de couples (Ip, Cp)** où Ip représente la p-ième classe d'indices à rechercher dans le contexte sémantique des arbres conceptuels Cp (partie « Conditions »), d'une **décision Dk** ainsi que d'un ensemble fini de triplets (S, L, T) où S représente la valeur de la décision Dk (partie « Action »), **L la position** de cette valeur dans le texte d'origine et T la valeur de confiance accordée à la décision (partie « Options »).

NomRègle: PersonnaliteR1	[nom de la règle]
TypeConcept: Classe Entité	
ConceptURI: <a href="http://www.mondeca.com/onto#Personnalité">http://www.mondeca.com/onto#Personnalité</a>	[partie Action]
NoeudIndicateur: NomPropre	[partie Déclenchement]
IndicesContextuels: Existe: [EspaceRechercheArbre: père] [NoeudIndice: Personne]	[partie Conditions]
Valeur: texte du nœud indicateur	[partie Options]
Position: faux	
Confiance: élevé	

**Figure 12 : Exemple d'une Règle d'Acquisition de Connaissance en langage OPAL** (Amardeilh, 2007).

La partie **Déclenchement** d'une RAC définit la classe N, la partie **Conditions** définit les couples (Ip, Cp) et la partie **Action** définit la décision Dk. Une RAC possède donc aussi une partie **Options**, ce en quoi elle se distingue des règles d'exploration contextuelle, partie Options qui définit le triplet (S, L, T). Nous allons à présent détailler chacun de ces composants.

#### 4.3.4.2 Les composants d'une RAC :

##### **La partie Déclenchement :**

Cette partie concerne l'information liée au déclenchement de la règle, i.e. l'étiquette sémantique qui va activer la règle. Cette étiquette sémantique est également appelée nœud indicateur ou nœud déclencheur. L'algorithme d'annotation et de peuplement d'ontologie d'OntoPop parcourt chaque nœud de l'arbre conceptuel généré par l'outil d'extraction d'information à la recherche du

prochain nœud porteur d'une étiquette sémantique paramétrée pour activer une Règle d'Acquisition de Connaissance.

### **La partie Conditions :**

La partie Conditions est optionnelle. Elle ne doit être déclarée que lorsqu'il est nécessaire de vérifier un ensemble de conditions auprès de certaines étiquettes sémantiques, i.e. les indices contextuels complémentaires, dans l'environnement du nœud déclencheur.

### **La partie Action :**

La partie Action est exécutée si et seulement si toutes les conditions de la partie précédente sont vérifiées. Cette partie précise l'élément de l'ontologie à instancier ou servant à créer l'annotation désirée ainsi que son type dans l'ontologie du domaine.

### **La partie Options : un apport des premiers projets.**

Au cours des premières expérimentations des RAC dans le cadre du projet Exinis (Amardeilh, 2004), il est apparu que les trois parties précédentes, bien qu'essentielles, n'étaient pas suffisantes pour capturer toute la complexité de la réalisation de la passerelle OntoPop. Les **RAC** ont été **complétées** par trois critères : la **valeur générée par la règle**, la **position de cette valeur** dans le document source et le **niveau de confiance** accordé à la règle.

Premièrement, la valeur générée par la règle en sortie ne correspond pas toujours à la valeur textuelle contenue dans le nœud indicateur, surtout lorsque l'application documentaire requiert une forte normalisation des connaissances extraites des documents (annotation de document juridique, par exemple). Il est donc parfois nécessaire de créer la valeur de l'instance ou de l'annotation en fonction de diverses informations comme le nom d'un nœud ou la concaténation de plusieurs valeurs de nœuds de l'arbre. Deuxièmement, il est intéressant de vouloir conserver les informations de position des informations extraites ou annotées dans le document d'origine, notamment à des fins de présentation pour l'utilisateur final. Troisièmement, chaque Règle d'Acquisition de Connaissance ne possède pas le même niveau de confiance. En effet, il existe deux sortes de méthodes permettant de repérer les entités nommées dans un document : soit à partir de lexiques contenant les entités nommées connues d'un domaine, soit à partir de patrons d'extraction qui détectent de nouvelles entités nommées potentielles. Ainsi, dans le cas d'une

application permettant de peupler une ontologie à partir de ces deux méthodes, il est important de savoir si la nouvelle instance d'entité nommée provient d'un lexique, auquel cas le taux de confiance dans l'information extraite est élevé, ou bien si elle a été déduite à partir d'un patron d'extraction, auquel cas le taux de confiance accordé sera plus faible. Par exemple, on dispose de deux sortes d'étiquettes sémantique dans notre application « People » : l'étiquette sémantique « /Personnalité » permet de savoir que la personnalité reconnue provient d'un lexique contrôlé et l'étiquette sémantique « /NomDePersonnePotentiel » que le nom de la personnalité a été déduite par un patron. Ainsi, deux Règles d'Acquisition de Connaissance sont définies à partir de ces étiquettes, l'une possédant un taux de confiance plus élevé que l'autre. Cette information peut être utilisée pour effectuer un nettoyage de la base de connaissance ou bien pour signaler à l'utilisateur le niveau de pertinence de l'information.

#### *4.3.5 Edition des Règles d'Acquisition de Connaissance grâce à un éditeur de règle :*

Dans la suite logicielle fournie par OntoPop, l'utilisateur dispose d'un logiciel Editeur de Règles qui lui permet de saisir des Règles d'Acquisition de Connaissance. Cet éditeur de règle n'avait toutefois pas encore atteint un degré de développement satisfaisant au terme du projet Exinis, pour pouvoir être utilisé par des non spécialistes, comme les clients (cf. partie 7 de ce chapitre). Son interface utilisateur propose un formulaire composé des champs nécessaires à la création d'une nouvelle règle comme spécifié par le langage OPAL. Ce langage original a été extrapolé du langage Longtext, spécialement pour le logiciel éditeur de RAC, par F. Amardeilh et du chef de produit « technique » d'ITM. Ensuite, les RAC sont traduites dans un langage informatique pour être comprises et manipulées par les agents logiciels. Lorsque l'utilisateur saisit une nouvelle règle d'acquisition dans l'Editeur des Règles, celle-ci est immédiatement compilée avec les règles déjà existantes pour générer deux feuilles de transformation au format XSLT.

#### *4.3.6 Conclusion :*

Les Règles d'acquisition de connaissances permettent de coupler un arbre conceptuel avec les éléments d'une ontologie. Ces règles présentent l'avantage de rendre la tâche de couplage théoriquement indépendante du choix des logiciels d'extraction d'information et de représentation des connaissances, même si elles n'avaient encore été appliquées qu'au couplage IDE/ITM, au terme de notre recherche.

## 5. FONCTIONNEMENT DE LA PASSERELLE ONTOPOP :

La démarche OntoPop repose sur les Règles d'Acquisition de Connaissance décrites précédemment, et sur les composants logiciels qui constituent cette passerelle. Ces traitements s'effectuent en fonction du cycle de vie des ressources terminologiques ou ontologiques (RTO, cf. point 6 de l'Annexe) dans OntoPop, c'est-à-dire les instances de l'ontologie et les descripteurs des thésaurus (cf. figure 13, page suivante).

### 5.1 Cycle vie des RTO dans la démarche OntoPop :

Pour présenter ces différentes composantes logicielles, et leur fonction dans l'application, nous allons suivre le traitement d'un document effectué par l'application de création automatisée de base de connaissance.

#### 5.1.1 *L'analyse linguistique :*

Le moteur d'extraction (IDE couplé à la cartouche de connaissance personnalisée) analyse le document soumis. Il repère les informations à extraire dans le document et les étiquette afin de générer en sortie un arbre conceptuel, en fonction de ses lexiques et de ses patrons d'extractions.

#### 5.1.2 *L'application des Règles d'Acquisition de Connaissance :*

La deuxième tâche est réalisée par le Module d'Annotation et d'Acquisition d'OntoPop qui se subdivise en deux composants : l'un dédié au peuplement de l'ontologie et l'autre dédié à l'annotation documentaire. Ce module dispose d'un ensemble de RAC définies préalablement pour l'application documentaire et les applique à l'arbre conceptuel du document analysé. Comme nous l'avons vu, les RAC sont déclenchées en fonction des nœuds indicateurs rencontrés dans l'arbre et de la résolution des indices contextuels relatifs à ces nœuds indicateurs.

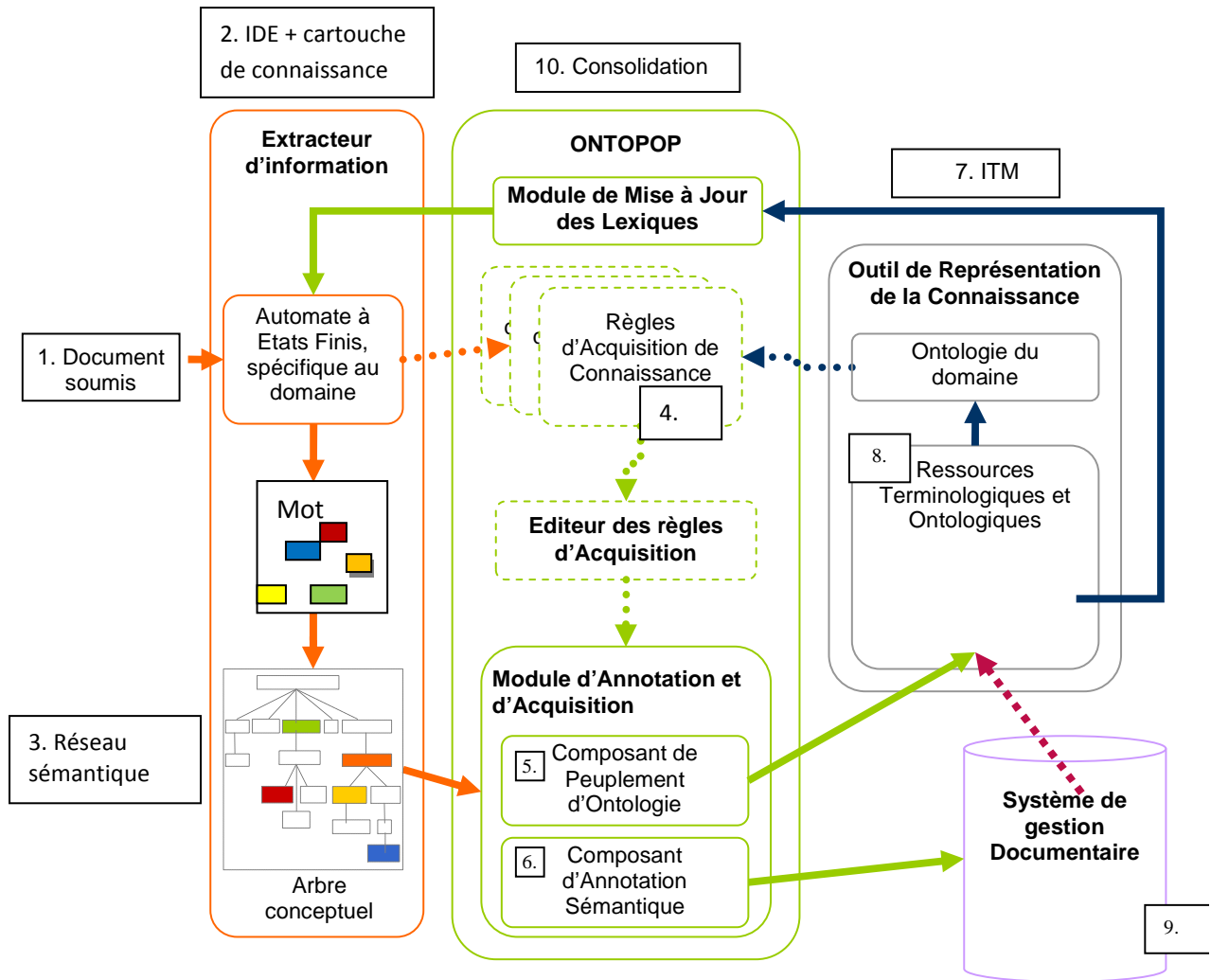


Figure 13 : Cycle de vie des RTO dans OntoPop. (D'après Amardeilh, 2007).

L'application de ces RAC produit deux sortes de résultats. Elles produisent un réseau sémantique de connaissance dit « potentiel » car en attente de consolidation par le Composant de Peuplement d'Ontologie. Elles produisent aussi un ensemble d'annotations sémantiques également « potentielles » devant être contrôlées par le Composant d'Annotation Sémantique. Chaque composant contrôle donc la validité des résultats produits ainsi que l'absence de redondance par rapport au référentiel terminologique et ontologique de l'outil de représentation des connaissances (ITM). Le réseau sémantique de connaissance est le premier à être contrôlé. Une fois le réseau sémantique contrôlé par le Composant de Peuplement d'Ontologie, il devient « valide » et il est importé dans l'outil de représentation des connaissances. De même, les annotations sémantiques contrôlées et « valides » vis-à-vis du Composant d'Annotation Documentaire sont mises à disposition d'un système de gestion de contenu documentaire.

### 5.1.3 *L'enrichissement des lexiques linguistiques de l'outil d'extraction de l'information via le module « mise à jour des lexiques » :*

Le module de mise à jour des lexiques a été conçu pour le projet Exinis. Il intervient lors de la troisième étape du cycle de vie des RTO dans OntoPop, et concerne l'enrichissement des lexiques de l'outil d'extraction d'information à partir des instances ou des descripteurs nouvellement créés dans l'outil de représentation de connaissance. Cette étape est réalisée par le Module de Mise à Jour des Lexiques. Il « écoute » en quelque sorte les créations, modifications ou suppressions qui sont opérées automatiquement par le Module d'Annotation et d'Acquisition, mais aussi manuellement par les utilisateurs humains de l'outil de représentation des connaissances. A chaque création, modification ou suppression d'une ressource terminologique ou ontologique servant à alimenter les lexiques de l'outil linguistique, le Module de Mise à Jour récupère une copie de cette ressource et la transfère à l'outil d'extraction d'information. Ce dernier peut alors créer une nouvelle entrée dans le lexique adéquat, ou bien mettre à jour l'entrée correspondante à cette ressource dans le cas d'une modification, ou encore supprimer cette entrée du lexique.

Nous allons à présent décrire le fonctionnement des deux modules logiciels qui fondent la passerelle OntoPop, i.e. le Module d'Annotation et d'Acquisition et le Module de Mise à Jour des Lexiques.

## 5.2 **Fonctionnement des modules logiciels d'OntoPop :**

### 5.2.1 *Le module d'annotation sémantique et de peuplement ontologique :*

Le processus d'annotation sémantique et de peuplement d'ontologie, opéré par le Module d'Annotation et d'Acquisition d'OntoPop, se décompose en trois grandes étapes :

1. **Transformer** les arbres conceptuels en des formats correspondant aux tâches de peuplement d'ontologie ou d'annotation sémantique ;
2. **Consolider** ces formats en fonction de la modélisation imposée par l'ontologie et du contenu du référentiel terminologique et ontologique ;

3. **Faire valider** les annotations ou les instances nouvellement créées par un utilisateur humain, dans le cas d'un processus semi-automatisé.

#### 5.2.1.1 La transformation :

Une fois l'ensemble des Règles d'Acquisition de Connaissance définies et compilées pour former les deux feuilles de transformation XSLT, le processus peut être déclenché sur un document ou un corpus de documents. Chaque document analysé produit un arbre conceptuel qui est « XML-isé » s'il n'a pas déjà été généré au format XML. Chacune des feuilles de transformation est appliquée à l'arbre conceptuel créant, d'une part un réseau sémantique de connaissance pour le peuplement de l'ontologie (au format XTM), et d'autre part un ensemble d'annotations sémantiques pour l'annotation documentaire (au format RDF).

La transformation XSLT correspond à une simple projection de l'arbre conceptuel vers un nouveau format. Cette projection induit un certain nombre de problèmes, problèmes mis en évidence dans le cadre du travail effectué par le projet PressPro, comme la redondance de l'information et la non-résolution des références directement liées à la transformation. En effet, si une entité nommée est citée plusieurs fois dans le document d'origine, elle est à chaque fois extraite par le moteur d'extraction et apparaît donc autant de fois dans l'arbre conceptuel. En conséquence, le réseau sémantique contient également autant d'instances identiques. Par ailleurs, la simple projection de l'arbre en réseau sémantique ne permet pas d'analyser les informations par rapport à un quelconque référentiel, ni même par rapport aux informations déjà présentes dans le réseau sémantique.

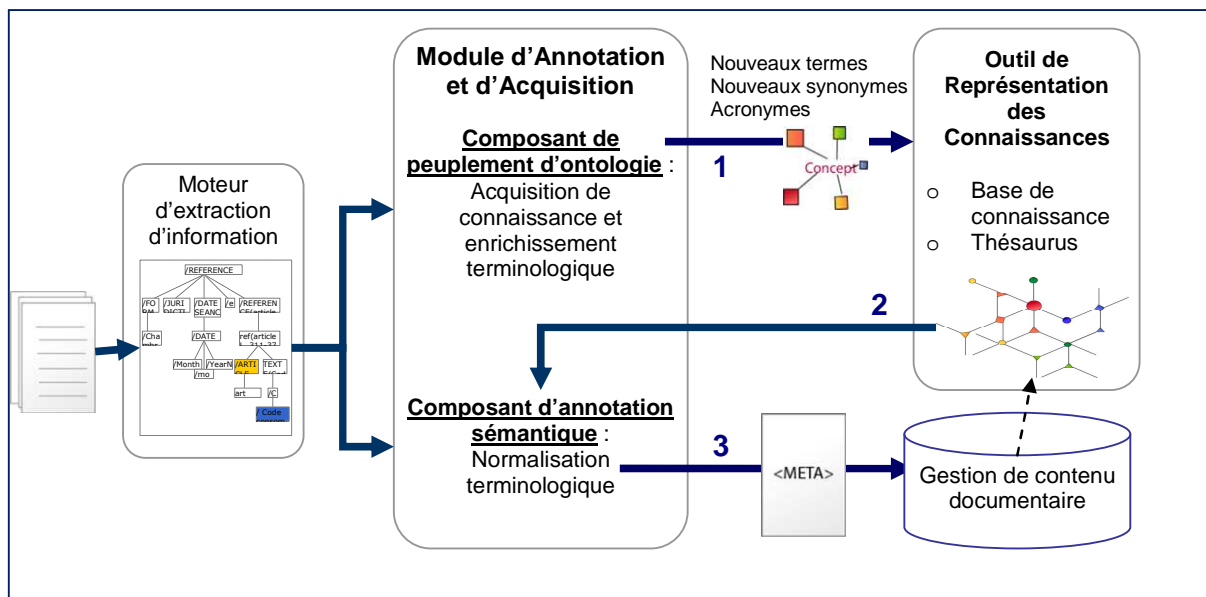
En conclusion, la projection de l'arbre conceptuel vers un nouveau format, quel qu'il soit, à partir d'une simple feuille de transformation XSLT, nécessite de retravailler les résultats obtenus. La résolution des références vis-à-vis du contenu du référentiel terminologique et ontologique est complétée par le contrôle du respect des contraintes imposées par la modélisation de l'ontologie du domaine. Il s'agit de l'opération de consolidation des informations afin de les rendre «valides». C'est l'objectif poursuivi par la prochaine étape du processus d'Annotation et d'Acquisition.

#### 5.2.1.2 La consolidation :

Cette étape est nécessaire au maintien de l'intégrité et de la qualité du référentiel de l'application. En fait, la plupart des outils font surtout appel à une validation manuelle pour vérifier les annotations ou les instances générées par leur système.

## Les Tests de consolidation proposés par OntoPop :

Cette étape de consolidation est réalisée avant la création des instances dans le référentiel. Cette opération s'effectue donc à partir des résultats obtenus à l'étape de transformation (figure 14). Ces résultats sont analysés en détail pour lever toute ambiguïté, toute inconsistance ou tout conflit avec des informations déjà existantes (tableau 2 « Opérations de consolidation réalisées par OntoPop en fonction des deux axes traités », page suivante).



**Figure 14 : Processus de consolidation des informations extraites (instances et annotations)** (Amardeilh, 2007).

En plus de ces opérations de consolidation, il faut procéder à l'évaluation du niveau de confiance accordée à chaque nouvelle instance ou annotation. Si deux règles concurrentes se déclenchent sur deux étiquettes ayant des valeurs différentes pour instancier le même élément, seule celle possédant le niveau de confiance le plus élevé sera conservée. L'autre sera supprimée. Ceci constitue le seul cas où de l'information est supprimée. Dans tous les autres cas de consolidation cités, si l'instance, le descripteur ou l'annotation est rejeté par l'étape de consolidation, il est conservé dans un «tampon» afin d'être ultérieurement proposé à l'utilisateur pour correction et validation (cf. Tableau 2, ci-après).

Éléments vs. Contraintes	Instance de classe entité	Instance d'attribut	Instance de classe relation	Descripteur candidat	Annotation sémantique
<b>Information dupliquée</b>	<p>Contrôle de l'existence de l'instance dans la base de connaissance par :</p> <ul style="list-style-type: none"> <li>→ recherche de son libellé ou de ses alias</li> <li>→ recherche de ses propriétés identifiantes (attributs obligatoires, i.e. dont les cardinalités sont non nulles)</li> </ul>	<p>Contrôle de l'existence d'un attribut pour une instance donnée par :</p> <ul style="list-style-type: none"> <li>→ recherche de ce type d'attribut sur l'instance donnée et vérification de sa valeur</li> </ul>	<p>Contrôle de l'existence d'une relation entre des instances données par :</p> <ul style="list-style-type: none"> <li>→ recherche de ce type de relation sur chacune des instances données et vérification de leurs valeurs</li> </ul>	<p>Contrôle de l'existence du descripteur dans un thésaurus de l'application par :</p> <ul style="list-style-type: none"> <li>→ recherche de son libellé ou de ses synonymes, variantes orthographiques ou traductions</li> </ul>	<p>Contrôle de l'existence d'une annotation liée à un document donné par :</p> <ul style="list-style-type: none"> <li>→ recherche de ce type d'annotation attaché au document donné et vérification de sa valeur (texte ou référence)</li> </ul>
<b>Domaine ou Classe</b>	<p>Contrôle de l'appartenance de l'instance à la bonne classe de l'ontologie ou à l'une de ses sous-classes</p>	<p>Contrôle de l'adéquation de la classe de l'instance à laquelle cet attribut est rattaché avec la modélisation de son domaine dans l'ontologie</p>	<p>Contrôle de l'adéquation de la classe de l'instance à laquelle cette relation est rattachée avec la modélisation de son domaine dans l'ontologie</p>	<p>Pas de contrôle, nouveau descripteur ajouté par défaut dans la classe « Descripteurs candidats »</p>	<p>Contrôle de l'adéquation de la classe de l'instance à laquelle cette annotation est rattachée avec la modélisation de son domaine dans l'ontologie</p>
<b>Portée ou Format de Données</b>	<p>Pas de contrôle</p>	<p>Contrôle de l'adéquation de la valeur de cet attribut avec la modélisation de son format de données dans l'ontologie (texte, date, numérique, etc.)</p>	<p>Contrôle de l'adéquation de la valeur (référence) de cette relation avec la modélisation de sa portée dans l'ontologie (classe autorisée et ses sous-classes)</p>	<p>Pas de contrôle</p>	<p>Contrôle de l'adéquation de la valeur (référence) de cette annotation avec la modélisation de sa portée dans l'ontologie (classe autorisée et ses sous-classes)</p>
<b>Cardinalité</b>	<p>Pas de contrôle</p>	<p>Contrôle du nombre d'attribut de ce type existants sur l'instance donnée</p>	<p>Contrôle de l'arité de la relation : les relations unaires ne sont pas considérées comme une vraie relation</p>	<p>Pas de contrôle</p>	<p>Pas de contrôle</p>

Tableau 2 : Opérations de consolidation réalisées par OntoPop en fonction des deux axes traités. (Amardeilh, 2007).

Enfin, lorsqu'un nouveau descripteur doit être créé dans un des thésaurus de l'application, le processus de consolidation ne peut savoir avec précision à quel emplacement du thésaurus concerné ce nouveau descripteur doit être enregistré : quelle sera sa hiérarchie parmi la taxonomie des descripteurs existants ? Quels seront ses termes associés ? Etc. Par conséquent, tout nouveau descripteur est enregistré dans une classe nommée « Descripteur Candidat » du thésaurus concerné. Son véritable emplacement est ensuite confirmé par un utilisateur.

En résumé, l'étape de consolidation implémentée dans le Module d'Annotation et d'Acquisition d'OntoPop consiste à :

- Contrôler les instances du réseau sémantique puis les annotations sémantiques en fonction du modèle de l'ontologie (restrictions de domaine, de couverture, de cardinalité), des vocabulaires contrôlés, comme les thésaurus ou les tables de références, et de la base de connaissance ;
- Importer les nouvelles instances valides dans la base de connaissance ou dans les vocabulaires contrôlés et créer les annotations sémantiques valides ;
- Enregistrer les instances et les annotations inconsistantes dans un tampon pour une validation manuelle de l'utilisateur final ultérieure.

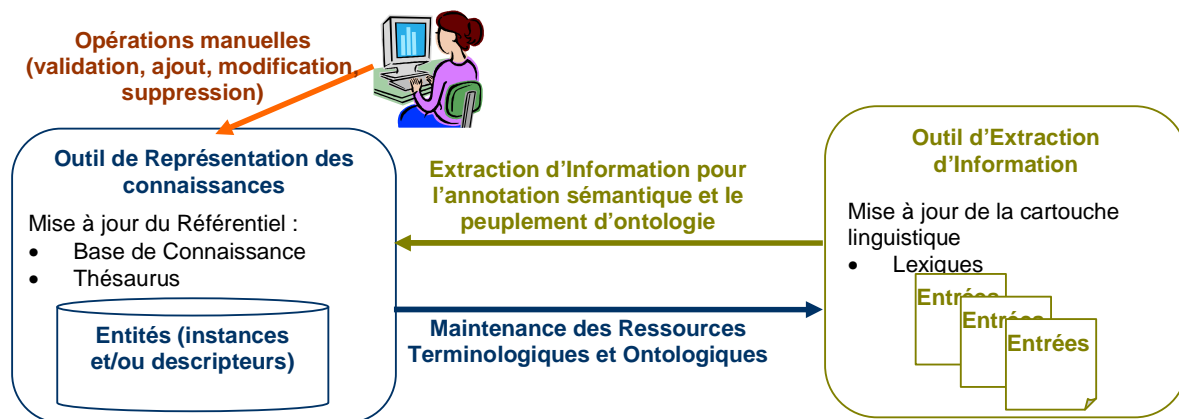
### 5.2.1.3 La validation :

Dans le cas d'une application semi-automatisée, l'utilisateur doit valider les résultats générés par le Module d'Annotation et d'Acquisition afin d'en contrôler la qualité. Pour cela, il dispose dans OntoPop, d'une interface unique, pour valider à la fois les annotations sémantiques et les instances créées. Il peut les éditer, les modifier ou encore les supprimer.

### 5.2.2 *La maintenance des lexiques et autres ressources linguistiques :*

Au projet Exinis, il est apparu nécessaire de capitaliser sur ces corrections et validations entrées par l'utilisateur. Mondeca a choisi d'intégrer ces corrections et validations au processus grâce à la maintenance des ressources terminologiques et ontologiques, i.e. la mise à jour des lexiques, et autres ressources linguistiques, des outils d'extraction d'information en fonction des entités (instances de la base de connaissance ou descripteurs d'un thésaurus) validées, créées, modifiées ou supprimées dans le référentiel d'une application donnée.

Cette maintenance s'applique donc aux nouvelles entités (instances et descripteurs) créées et validées dans le référentiel dans le cadre de la tâche de peuplement de l'ontologie. Dans le cas de la connaissance déduite, le référentiel peut ne pas posséder d'entité correspondant à cette nouvelle connaissance. C'est cette nouvelle connaissance qui est exportée vers l'outil d'extraction d'information afin qu'elle soit intégrée à ses lexiques puis reconnue. Comme illustré par la figure 15, ces entités enrichissent les lexiques de la cartouche linguistique conçue pour le domaine de l'application.



**Figure 15 : Opération de consolidation des informations (instances et annotations) extraites** (Amardeilh, 2007).

La mise à jour des lexiques du moteur d'extraction est déclenchée en fonction d'un ensemble d'alertes configurées dans l'outil de représentation des connaissances sur chacune des classes du référentiel dont les instances doivent être exportées. Ces alertes « écoutent » les différentes opérations effectuées sur ces entités. En synthèse, chacune des entités exportées est analysée par le moteur d'extraction qui complète ses dictionnaires d'entités nommées ou toute autre ressource lexicale avec les différents termes attachés à l'entité. Ces nouvelles ressources linguistiques sont ensuite compilées dans une nouvelle version de la cartouche linguistique de l'application afin d'être prises en compte durant la prochaine tâche d'extraction d'information.

### 5.3 Les composants de la passerelle logicielle OntoPop :

En conclusion, c'est le projet PressPro qui a permis la conception de la passerelle logicielle OntoPop. Cette passerelle est constituée de différents composants logiciels modulaires qui permettent de réaliser des applications destinées au peuplement d'ontologie et à l'annotation sémantique, ce que nous appelons la « création automatisée de base de connaissances ». L'architecture de cette passerelle OntoPop a d'abord été conçue au cours du projet PressPro,

autour de deux composants logiciels : l'éditeur de règles d'acquisition de connaissance et le module d'annotation et d'acquisition. Ensuite, le projet Exinis (cf. partie 7 de ce chapitre), a conduit à la définition d'un composant logiciel supplémentaire, le module de maintenance des lexiques, ainsi qu'à l'enrichissement du concept de RAC, au travers de l'ajout d'une partie « Options ». Ce composant supplémentaire et cette partie additionnelle ont alors été intégrés de manière générique à la passerelle OntoPop. Plus globalement, le travail sur chacun des composants de cette passerelle s'est poursuivi et affiné au cours du projet Exinis, après sa définition originelle dans le cadre du projet PressPro. De plus, la mise en œuvre d'une solution de création automatisée de base de connaissance, fondée sur les logiciels IDE et ITM, et sur la passerelle OntoPop, a nécessité la conception d'une méthodologie que nous décrivons maintenant.

## **6. LA METHODOLOGIE ONTOPOP.**

### **6.1. Introduction :**

Les projets PressPro et Exinis ont donné lieu à une réflexion sur la meilleure manière de mener à bien un projet de création automatisée de base de connaissance. A cette fin, Mondeca, par l'intermédiaire de la responsable de couplage en collaboration avec l'expert ontologue et le chef de projet de Mondeca, ainsi qu'avec le chef de projet de Temis, a progressivement élaboré une méthodologie relative, principalement, au couplage entre moteur d'extraction d'information et système de représentation des connaissances. L'élaboration de cette méthodologie, mise en œuvre à partir de juillet 2006, visait à résoudre des problèmes ressentis par les acteurs des projets PressPro et Exinis, entre 2003 et 2006, problèmes qui ont perturbé leur travail. Ces acteurs ont en effet été confrontés à un certain nombre de questions et d'incertitudes, qui ont été source d'incompréhensions ou de pertes de temps. Ainsi, par exemple, l'interdépendance des tâches nécessitent qu'à certains moments du processus, un ou plusieurs acteur(s) suspende(nt) son [leur] travail (phase appelée de « stabilisation »), pour permettre à un ou plusieurs autre(s) acteur(s) de travailler sur ce que ce qui a été « stabilisé ». Des questions ont donc émergé, parmi lesquelles : Quelles sont les différentes phases du processus de réalisation de la solution ? Comment coordonner les tâches de développement de cartouche de connaissance, de conception d'ontologie et de définition des RAC ? Quels doivent être les « livrables » intermédiaires ?

## 6.2 Présentation générale de la méthodologie OntoPop :

Depuis le début de sa thèse en novembre 2003, F. Amardeilh a participé aux projets PressPro et Exinis en tant que responsable de couplage et réalisé les couplages nécessaires au développement soit de prototypes de démonstration de la solution, soit de la future application mise en service. Elle a conçu la méthodologie OntoPop grâce à l'expérience acquise au cours des projets PressPro et de la première partie du projet Exinis, i.e. jusqu'au développement de la version V0. Elle s'est inspirée des méthodes de gestion de projet, mais aussi et surtout des réussites et des dysfonctionnements apparus au fil de ces projets. Nous présentons ici de manière synthétique la méthodologie OntoPop, telle qu'elle a été définie et validée par Mondeca et Temis. Les terminologies que nous présentons ci-après (partie 6 du chapitre 4), en particulier celles qui concernent les acteurs, sont celles employées par Temis et Mondeca.

La méthodologie OntoPop consiste tout d'abord à attribuer des rôles aux différents acteurs du projet, et à décrire leur collaboration (réunions de travail, « livrables » à fournir). , au cours des cinq phases successives du projet : étude, structuration, couplage, validation/qualité, livraison. Cette méthodologie spécifie différents rôles :

- **L' « expert du domaine »** ou simplement « expert » (a priori un acteur situé chez le client) décrit les besoins des utilisateurs, spécifie les limites de l'application et du domaine concerné, fournit les ressources existantes et enfin valide l'ensemble de la solution.
- **L' « expert linguiste »**, ou simplement « linguiste », se charge de l'analyse textuelle en fonction d'un corpus de documents représentatifs et livre un outil d'extraction d'information adapté au domaine.
- **L' « expert ontologue »**, ou simplement « ontologue », voire « ontographe » modélise l'ontologie du domaine et reprend les données existantes (bases de données, thésaurus, terminologies, etc.) de l'application source pour les enregistrer dans la base de connaissance et les thésaurus de l'application documentaire.
- **Le « responsable de couplage »** ou chef de projet couplage (appelé en interne « intégrateur de couplage », à ne pas confondre avec l'intégrateur de la solution complète) met en place les différents outils et implémente la solution dans sa globalité.

Le responsable de couplage doit modéliser les RAC, sur la base des informations fournies par le linguiste, le client et l'ontologue. Mais avant de pouvoir définir ces règles, il faut savoir évaluer les charges de spécifications et de développement inhérentes à chacun des outils utilisés. Il est pour cela nécessaire :

- a) d'estimer la capacité d'extraction d'information sur un nouveau domaine et la couverture pouvant être atteinte sur des corpus documentaires représentatifs, et
- b) de modéliser l'ontologie en tenant compte à la fois de la formalisation, de la couverture ciblée pour la nouvelle application et de la reprise de l'existant.

Les projets PressPro et Exinis ont montré qu'il est important de mettre en commun, aussi rapidement que possible dans le déroulement d'un projet, le besoin client et ses problématiques spécifiques. Il faut aussi identifier rapidement la couverture commune au moteur d'extraction et à l'ontologie du domaine. Pour ce faire, des documents d'échange, comme la structure des ressources linguistiques et de l'ontologie, doivent être partagés entre les différents intervenants afin de réaliser un premier couplage dans les plus courts délais. Ce dernier sera ensuite itérativement développé, testé et affiné avant d'être finalement validée par le client. Ces quelques principes de base ont permis de définir la méthodologie OntoPop, inspirée également des recommandations du génie logiciel (Schultz, 1997) autour des étapes d'évaluation des besoins, de construction de la solution logicielle, de sa validation et de son utilisation dans une application donnée. Le cycle de vie de cette méthodologie est le suivant :

1. **La Phase d'Etude** : Etude sur les données à traiter par l'application documentaire ainsi que leurs sources et définition de la couverture du domaine ;
2. **La Phase de Structuration** : Définition de la structure des arbres conceptuels, résultats des outils de TAL (traitement automatique du langage), et modélisation de l'ontologie du domaine ;
3. **La Phase de Couplage** : Couplage des arbres conceptuels avec les éléments de l'ontologie et définition des Règles d'Acquisition de Connaissance ;
4. **La Phase de Validation/Qualité** : Validation des annotations et des instances créées et réitération si besoin est à partir de la Phase de Structuration.
5. **La Phase de Livraison** : Livraison et maintenance auprès du client.

Chacune de ces étapes est décrite de manière synthétique ci-dessous. Ce tableau (tableau 3, page suivante) qui a été formalisé dans la thèse de F. Amardeilh (2007), soit en 2006, alors que le projet PressPro était terminé, et le projet Exinis encore en cours, est désormais un document de référence opérationnel à la fois chez Mondeca et chez Temis. Il est intéressant de noter comment de petites entreprises ayant peu de documents formalisant leur processus de travail ont pu solliciter des doctorants pour en structurer, à côté d'une contribution opérationnelle directe à des projets.

	Phase d'Etude	Phase de Structuration	Phase de Couplage	Phase de Validation / Qualité	Phase de Mise en Service
<b>Objectifs</b>	<ul style="list-style-type: none"> <li>- Etude sur les données à traiter par l'application cible ainsi que leurs sources</li> <li>- Définition de la couverture du domaine</li> </ul>	<ul style="list-style-type: none"> <li>- Mise en place de la structure de l'arbre conceptuel</li> <li>- Modélisation de l'ontologie du domaine</li> </ul>	<ul style="list-style-type: none"> <li>- Couplage des arbres conceptuels avec les éléments de l'ontologie</li> <li>- Définition et écriture des RAC</li> </ul>	<ul style="list-style-type: none"> <li>- Validation des résultats obtenus</li> <li>- Réitération à partir de la Phase de Couplage ou de Structuration selon les corrections nécessaires</li> </ul>	<ul style="list-style-type: none"> <li>- Livraison de l'application</li> <li>- Formation des utilisateurs</li> <li>- Maintenance de l'application</li> </ul>
<b>Actions</b>	<ul style="list-style-type: none"> <li>- Etude du domaine de l'application cible, de ses besoins et de l'existant</li> <li>- Etude des corpus documentaires</li> <li>- Délimite du périmètre de l'application</li> <li>- Délimite de la couverture des deux outils</li> </ul>	<ul style="list-style-type: none"> <li>- Etude fine des besoins de l'application cible et du contenu de ses sources de données</li> <li>- Interactions sur la future structure des arbres conceptuels</li> <li>- Interactions sur le futur modèle de l'ontologie</li> <li>- Analyse régulière de leurs correspondances</li> </ul>	<ul style="list-style-type: none"> <li>- Implémentation de l'ontologie dans l'outil de représentation</li> <li>- Rapprochement entre les étiquettes linguistiques et les éléments ontologiques</li> <li>- Analyse du contexte de chaque étiquette linguistique</li> <li>- Implémentation des RAC</li> </ul>	<ul style="list-style-type: none"> <li>- Constituer le corpus de test du domaine</li> <li>- Mesurer la qualité de la cartouche linguistique</li> <li>- Mesurer la qualité du couplage</li> <li>- Corriger les règles</li> <li>- Améliorer la cartouche et/ou l'ontologie</li> <li>- Valider les résultats</li> </ul>	<ul style="list-style-type: none"> <li>- Livraison et Installation des composants</li> <li>- Recette de l'application</li> <li>- Mise en production de l'application auprès des utilisateurs</li> </ul>
<b>Moyens</b>	<ul style="list-style-type: none"> <li>- Mise à disposition par l'expert des différentes sources de données existantes</li> <li>- Entretiens avec l'expert</li> <li>- Réunion entre les différents acteurs</li> </ul>	<ul style="list-style-type: none"> <li>- Echanges réguliers entre l'expert et l'ontologue et entre l'expert et le linguiste</li> <li>- Réunion de validation des modèles avec en plus le responsable de couplage</li> </ul>	<ul style="list-style-type: none"> <li>- Analyse fine des arbres conceptuels</li> <li>- Réunions de travail entre le linguiste et le responsable de couplage.</li> </ul>	<ul style="list-style-type: none"> <li>- Réunions de travail entre le responsable de couplage et le linguiste pour confronter les résultats obtenus et décider des modifications</li> </ul>	<ul style="list-style-type: none"> <li>- Réunion de validation entre tous les intervenants</li> <li>- Formation des utilisateurs finaux et de l'administrateur système</li> </ul>
<b>Livrables</b>	<ul style="list-style-type: none"> <li>- Document de spécifications générales de l'application cible</li> <li>- Structure des sources de données existantes : bases de données, bases de connaissances, thésaurus, etc.</li> <li>- Corpus de documents textuels représentatif du domaine de l'application</li> </ul>	<ul style="list-style-type: none"> <li>- Ontologie du domaine avec sa documentation</li> <li>- Cartouche linguistique avec les spécifications détaillées de la structure des arbres conceptuels du domaine</li> <li>- Document « Table de correspondance cartouche/ontologie »</li> </ul>	<ul style="list-style-type: none"> <li>- Nouvelle version de l'ontologie du domaine, et sa documentation</li> <li>- Nouvelle version de la cartouche linguistique et sa documentation</li> <li>- La « Table de correspondance cartouche/ontologie » complétée des règles d'acquisition de connaissance</li> </ul>	<ul style="list-style-type: none"> <li>- Nouvelle version de l'ontologie du domaine</li> <li>- Nouvelle version de la cartouche linguistique</li> <li>- Le couplage validé</li> <li>- Le rapport de test du couplage pour chaque règle et le rapport de couverture entre les deux outils</li> <li>- La « Table de correspondance cartouche/ontologie » mise à jour avec éléments validés</li> </ul>	<ul style="list-style-type: none"> <li>- Les deux outils (si non déjà installés)</li> <li>- L'ontologie du domaine finalisée + documentation</li> <li>- La cartouche linguistique finalisée + documentation</li> <li>- Le couplage finalisé</li> <li>- La « Table de correspondance cartouche/ontologie » mise à jour</li> <li>- Les rapports de tests de qualité (recette)</li> </ul>

Tableau 3 : Les étapes de la « méthodologie OntoPop » (d'après Amardeilh, 2007).

## 6.3 Les différentes phases de la méthodologie OntoPop :

### 6.3.1 La Phase d'Etude :

Cette phase consiste à déterminer les matériaux de base nécessaires à l'analyse linguistique mais aussi à la représentation des connaissances destinées à l'application documentaire. Au cours de cette étape de familiarisation, l'ontologue et le linguiste cherchent à comprendre à la fois le domaine concerné, le vocabulaire utilisé pour décrire ce domaine et les problématiques que la future application doit résoudre. L'expert leur fournit le support nécessaire à la compréhension

des points clefs du domaine, de ses frontières ainsi que des objectifs de l'application. Il doit pour cela formuler des réponses claires et précises à un ensemble de questions, parmi lesquelles :

1. Des questions d'ordre général, comme :

- L'application est-elle axée sur le peuplement d'ontologie, l'annotation documentaire ou les deux ?
- L'application doit-elle être entièrement automatisée ou semi-automatisée avec validation des résultats par un utilisateur final ?
- Existe-t-il des spécifications fonctionnelles et techniques de l'application documentaire ?
- Quels sont les points sensibles dans les processus de l'application existante devant être résolus dans l'application documentaire ?

2. Des questions d'ordre documentaire, comme :

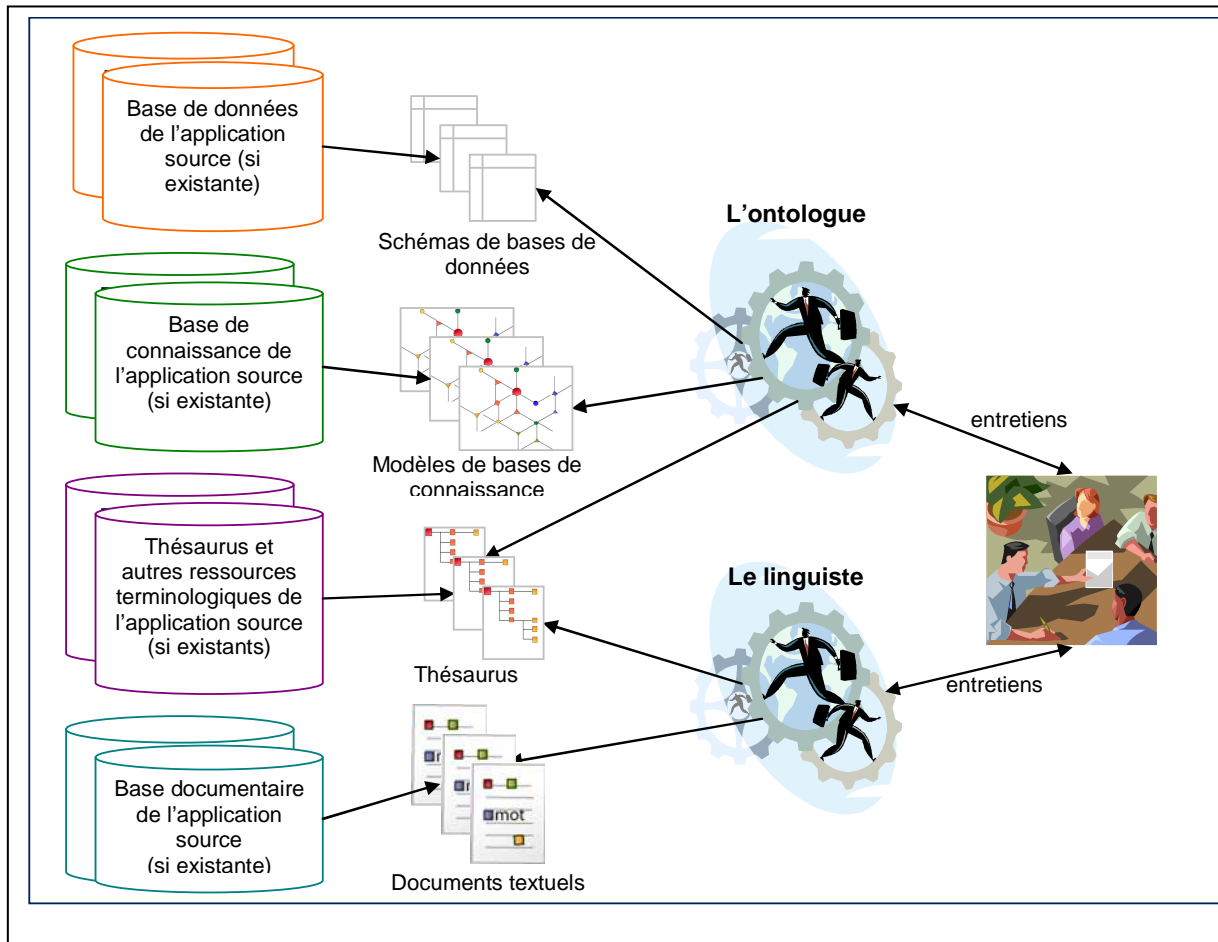
- Quels sont les flux documentaires existants et à prévoir dans la future application ?
- Quels sont les documents textuels ou autres représentatifs du domaine ?
- Quels sont les documents textuels manipulés par l'application existante et devant l'être par l'application documentaire ?
- Quels sont les formats des ressources textuelles (Word, PDF, XML, HTML, etc.) ?

3. Des questions concernant la gestion des connaissances, comme :

- Existe-t-il des bases de données, ou mieux, des bases de connaissances, liées à l'application existante et devant être reprises dans l'application documentaire ?
- Existe-t-il des ressources terminologiques, sous la forme de lexiques, de taxonomies ou de thésaurus, correspondant au domaine concerné ?
- Les données existantes doivent-elles être conservées dans la future application ?
- Leur structure (schéma ou modèle) doit-elle être également préservée ou est-il possible de repenser complètement cette structure ?
- Les documents déjà annotés, s'il existe une base documentaire existante, doivent-ils être réannotés en fonction du nouveau modèle de représentation des connaissances ?

En fonction des données existantes, le linguiste et l'ontographe vont étudier le domaine afin de cerner les possibilités tant au niveau de la couverture du moteur d'extraction sur le corpus documentaire qu'au niveau de la modélisation d'une ontologie de domaine (Figure 16, page suivante).

Le linguiste procède par une approche en mode découverte (Enjalbert 2005), c'est-à-dire qu'il part des données textuelles présentes dans les documents du corpus représentatif fourni par l'expert afin d'évaluer le degré d'analyse linguistique nécessaire. Il se familiarise avec le vocabulaire utilisé, avec la structure des phrases mais aussi avec la structure même des documents qui peuvent aussi apporter des éléments importants pour la construction des patrons d'extraction. En ce qui concerne l'élaboration des ressources linguistiques, et notamment des différents lexiques utilisés par les patrons d'extraction, il va recueillir et analyser toutes les ressources terminologiques mises à sa disposition par l'expert. Ces ressources terminologiques comprennent aussi bien des lexiques, des taxonomies, des thesaurus que des listes d'entités nommées exportées des bases de données ou des bases de connaissances existantes. Enfin, afin de cerner la couverture de la cartouche linguistique du domaine et les informations essentielles à extraire pour l'application finale, le linguiste procède en plus à une série d'entretiens avec l'expert. Celui-ci va notamment détailler la manière dont les informations extraites vont être exploitées par l'application documentaire. Il peut aussi lui préciser les besoins spécifiques à l'application documentaire, comme la normalisation des informations extraites.



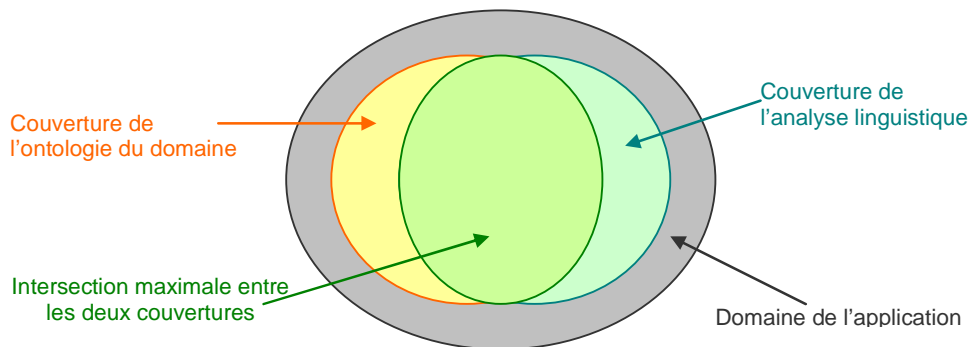
**Figure 16 : Sources des données pour le linguiste et l'ontologue dans la phase d'étude** (d'après Amardeilh, 2007).

Concernant la future ontologie, l'ontologue procède par une approche « métier », c'est-à-dire qu'il va partir de l'étude de l'existant (tant au niveau des bases de données, que des bases de connaissances ou encore des ressources terminologiques), des processus mis en place et des habitudes métier des utilisateurs. Il cible avec l'expert les spécificités métier, les référentiels à intégrer, les concepts clefs de l'application et plus généralement du domaine, ainsi que leurs attributs et leurs interactions. Il analyse également les schémas des bases de données et les modèles des bases de connaissances à reprendre, lorsqu'elles existent. Il peut aussi s'inspirer de sources de connaissances externes correspondant au domaine concerné par la future application, comme l'intégration d'un thésaurus géographique.

Cette étape est validée par la confrontation des différents intervenants afin que le linguiste et l'ontologue puissent divulguer les conclusions de leurs études et surtout, expliquer à l'expert ce qui est réalisable ou non. Chacun participe alors à l'écriture des spécifications générales qui

permettent d'une part de délimiter le périmètre de l'application finale et d'autre part de se mettre d'accord sur les objectifs à remplir concernant l'application finale. Ces spécifications générales valident aussi les différents cas suivants :

- L'outil linguistique extrait toutes les informations nécessaires au peuplement de l'ontologie et/ou à l'annotation sémantique
- L'outil de représentation des connaissances modélise le domaine de manière à ce que chaque information extraite par l'outil linguistique puisse être enregistrée dans la base de connaissance ou servir à créer de nouvelles annotations documentaires.



**Figure 17 : Comparaison de la couverture de l'ontologie du domaine et de la couverture de l'analyse linguistique** (Amardeilh, 2007).

Il est rare que la couverture de l'outil de représentation des connaissances vis-à-vis du domaine soit exactement égale à la couverture de l'outil d'analyse linguistique pour le même domaine (figure 17). En effet, les noms des futurs concepts de l'ontologie ne correspondent pas forcément aux labels des futures étiquettes linguistiques. Certains éléments de l'ontologie sont destinés à d'autres usages métiers de l'application finale (recherche d'information, publication, autres fonctionnalités). Et la réutilisation de cartouches linguistiques génériques, comme celle des entités nommées, ou la nécessité de construire des patrons d'extraction intermédiaires, créent, dans l'arbre conceptuel généré, des étiquettes linguistiques dont les valeurs ne sont pas forcément exploitées ni exploitables pour le peuplement d'ontologie ou l'annotation sémantique. Malgré tout, il est fortement souhaitable pour l'application documentaire que l'intersection entre ces deux couvertures soit maximale, avec le plus de précision et de pertinence possible, ce qui constitue l'objectif de la phase suivante.

### 6.3.2. La Phase de Structuration :

Cette phase a pour objectif d'une part de modéliser l'ontologie du domaine et d'autre part de construire la cartouche linguistique appliquée à ce domaine.

#### 6.3.2.1 Modélisation de l'ontologie du domaine :

Cette tâche consiste à structurer et organiser les données obtenues à l'étape précédente. Pour cela, l'ontologue doit acquérir un certain niveau de connaissance du domaine, que cette connaissance soit explicite dans les bases de données ou les thésaurus disponibles ou implicite dans le savoir et le savoir-faire de l'expert du domaine. En fait, il n'existe pas de méthodologie générique, standardisée et consensuelle ayant pour but d'acquérir la connaissance d'un domaine afin de la modéliser sous la forme d'une ontologie. Plusieurs méthodologies ont vu le jour, notamment celles formulées par Uschold et King (1995), Gruninger et Fox (1995), ou Fernandez & Gomez-Perez dans Methontology (Fernandez et al. 1997). Le cycle de vie de ces méthodologies est fortement inspiré du génie logiciel, tout comme la méthodologie OntoPop, et nous pouvons identifier des étapes communes telles que :

- la spécification / l'évaluation du besoin
- la conceptualisation, i.e. la capture des connaissances
- la formalisation ou « l'ontologisation », i.e. le codage de l'ontologie
- l'intégration d'ontologies existantes, par alignement ou par fusion entre ces ontologies
- l'opérationnalisation, i.e. l'implémentation de l'ontologie
- l'évaluation, la documentation et la maintenance de l'ontologie

Diverses méthodes, techniques et outils ont été proposés pour réaliser ces différentes tâches, et notamment au niveau de la conceptualisation, comme l'extraction d'ontologies à partir de textes (Bourigault et al. 2004), la structuration des hiérarchies de concepts et de relations (Guarino 1992), la fusion et l'adaptation d'ontologies existantes par l'utilisation des systèmes Onions et Prompt par exemple, le développement collaboratif (Domingue, 1998). Mais quelle que soit l'approche adoptée, le processus de construction d'une ontologie reste basé sur une collaboration étroite avec l'expert du domaine qui doit valider le modèle choisi.

### 6.3.2.2 Construction des cartouches linguistiques :

Concernant la construction des cartouches linguistiques, celles-ci s'appuient le plus souvent sur différents lexiques du domaine ainsi que sur un ensemble de patrons d'extraction composants l'automate à états finis du moteur d'extraction. Pour déterminer les patrons d'extraction, le linguiste étudie minutieusement chaque document du corpus représentatif du domaine, afin d'identifier le vocabulaire spécifique utilisé par les auteurs de ce domaine, la structure des phrases, celle des documents, etc. (Enjalbert 2005). Le contenu d'une décision de jurisprudence dans le domaine juridique est, par exemple, très différent de celui d'un article journalistique du domaine de la presse « people » :

- Le vocabulaire de la décision est constitué de termes juridiques très précis alors que celui de la presse « people » est constitué de termes du langage commun, voire familier ;
- Une décision de jurisprudence est souvent décrite par une seule phrase de plusieurs pages alors que l'article « people » se compose de phrases simples et courtes. Le document représentant la décision de jurisprudence comporte une structure bien particulière avec un en-tête composé de tous les éléments identificatoires de cette décision (cour de justice, date de la décision, la juridiction, etc.) suivi d'un corps de document narratif des arguments des différentes parties jusqu'au rendu de décision par la cour. Dans ce cas, le linguiste peut plus facilement cerner l'emplacement d'une information à extraire et la manière dont elle peut être extraite, que dans le cas d'un texte moins structuré comme un article de la presse « people » qui n'est généralement constitué que d'un titre, du corps de l'article et parfois d'un chapeau

Au vu du corpus documentaire du domaine, l'une des premières tâches du linguiste consiste à élaborer les lexiques de ce domaine nécessaires à l'élaboration des patrons d'extraction. Dans certains cas, il aura à sa disposition des thésaurus, des listes d'entités nommées, ou toute autre terminologie, déjà exploités dans une version précédente de l'application et fournis par l'expert, ou bien qui font référence dans le domaine concerné. Une autre étape consiste à définir les différents traitements linguistiques nécessaires à l'analyse des documents du corpus : est-il nécessaire de découper le document en unités textuelles ? Ces unités textuelles sont-elles les

paragraphes, les phrases, les mots ou autres ? L'analyse morphosyntaxique est-elle absolument requise (cf. Annexes, p. 355) ? Et ainsi de suite... La décision d'inclure tel ou tel traitement dans la solution finale du moteur d'extraction est généralement dictée par la langue dans laquelle est rédigé le document, son niveau de structuration, la nature des informations à extraire, etc.

<p>a) « isManagerOf »</p> <p><b>Objectif</b> Relation entre une organisation (reconnue ou déduite) et une personne (reconnue ou déduite).</p> <p><b>Lexique utilisé</b> Le lien « isManagerOf » (concept au sein de la <b>Skill Cartridge™</b>) s'effectue au moyen d'un lexique existant (Module NER) contenant des termes du type « <i>manager, pdg, président directeur général</i>, etc. ».</p> <p><b>Implémentation</b> La personne identifiée par cette relation porte l'étiquette « Manager » (rôle au sein de la <b>Skill Cartridge™</b>). La relation elle-même porte l'étiquette « isManagerOf » (concept au sein de la <b>Skill Cartridge™</b>).</p> <p><b>Exemples d'extraction</b></p> <div style="border: 1px solid black; padding: 5px;"> <p><i>Monsieur Thierry, pdg des Agf.</i></p> <p>isManagerOf → <i>Monsieur Thierry, pdg des Agf.</i>            Manager → <i>Monsieur Thierry</i>                GUESSED Personnalite → <i>Monsieur Thierry</i>                Title → <i>Monsieur</i>                FamilyName → <i>Thierry</i></p> <p>Board → <i>pdg</i>            Organisation → <i>Agf</i></p> </div> <div style="border: 1px solid black; padding: 5px; margin-top: 5px;"> <p><i>Olivier Maschino PDG de la société Carlogo.</i></p> <p>isManagerOf → <i>Olivier Maschino PDG de la société Carlogo.</i>            Manager → <i>Olivier Maschino</i>                GUESSED Personnalite → <i>Olivier Maschino</i>                FirstName → <i>Olivier</i>                FamilyName → <i>Maschino</i></p> <p>Board → <i>PDG</i>            GUESSED Organisation → <i>la société Carlogo</i></p> </div>
---

**Figure 18 : Extrait d'un document de spécifications détaillées des arbres conceptuels délivré par un linguiste de Temis pour une application d'intelligence économique, l'exemple est « isManagerOf » (Amardeilh, 2007).**

Une fois les traitements définis, il reste à écrire les patrons d'extraction. Pour cela, il est parfois possible de réutiliser certaines cartouches linguistiques comportant des patrons d'extraction génériques pour certains éléments particuliers à un langage. L'exemple le plus courant est celui des entités nommées comme les noms de personnes, d'organisations, de lieux, etc. Par exemple, dans le domaine de la presse « people », une cartouche existante permettant d'extraire le nom des personnes et donc des personnalités en vogue, peut être réutilisée telle quelle. Par contre, dans le

domaine juridique, cette même cartouche linguistique n'est pas suffisante pour décrire les différentes personnes intervenant dans une décision de jurisprudence : le nom des parties, celui des avocats, du juge, du président, etc.

Cette cartouche doit être adaptée en lui ajoutant de nouveaux patrons d'extraction permettant de repérer et de distinguer clairement ces différents acteurs. Le linguiste étudie donc dans quelle mesure il est possible de réexploiter des cartouches existantes. A partir de là, il lui faut créer les patrons d'extraction spécifiques au domaine de l'application documentaire qui viennent enrichir et compléter les patrons existants le cas échéant.

Tout au long de ce processus, le linguiste est en contact permanent avec l'expert pour lui présenter ses résultats, discuter des points à améliorer, etc. L'expert valide les informations extraites par le moteur d'extraction à partir d'un document des spécifications détaillées de la cartouche linguistique remis par le linguiste. Ce document détaille la structure de chaque arbre conceptuel généré à partir d'un patron d'extraction du domaine. La figure 18 (page précédente) montre un exemple du contenu d'un tel document pour un sous-arbre en particulier, i.e. « isManagerOf », issu d'une cartouche sur le domaine de la veille économique. Ce document doit être régulièrement mis à jour en fonction des modifications opérées sur la structure des patrons. A chaque nouvelle version de la cartouche linguistique, il est de nouveau livré à l'expert pour validation. Il est aussi fourni au responsable de couplage qui l'utilise comme base de travail pour définir les Règles d'Acquisition de Connaissance dans la prochaine étape.

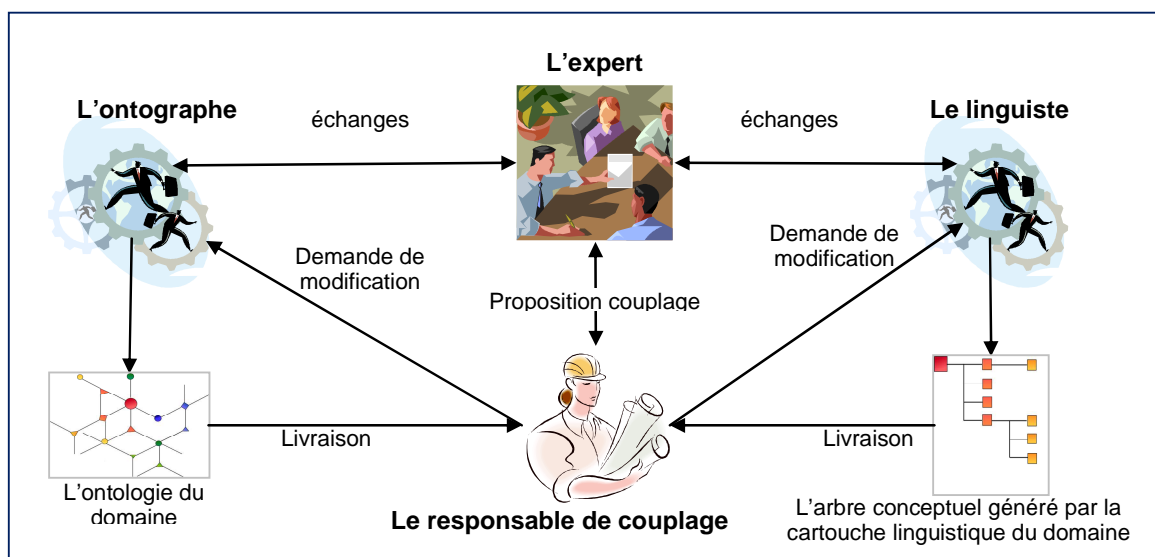


Figure 19 : Echanges entre les intervenants pendant la phase de Structuration (d'après Amardeilh, 2007).

A ce stade de la méthodologie, le responsable de couplage est responsable de la compatibilité des deux modèles vis-à-vis du domaine, et plus particulièrement des objectifs décrits dans les spécifications générales de l'application documentaire. Dès les premières versions, il initie le couplage entre les concepts de l'ontologie d'un côté et les étiquettes sémantiques de l'arbre conceptuel généré par la cartouche linguistique de l'autre. Cette mise en correspondance lui permet de détecter en amont les failles et de demander les modifications nécessaires au linguiste, à l'ontologue, ou au deux si nécessaire. Le responsable de couplage joue donc un rôle de médiateur entre les différents intervenants (figure 19), et notamment auprès de l'expert. Même si l'expert possède une certaine connaissance du domaine et de l'application documentaire, a priori, il ne maîtrise pas assez le fonctionnement des outils d'extraction d'information ou de représentation des connaissances, leurs formats, leurs contraintes, etc. Seul le responsable de couplage possède ce double point de vue et par conséquent les capacités de réaction associées.

Afin de rendre ce double point de vue accessible aux autres intervenants et tout particulièrement à l'expert, le responsable de couplage rédige un document nommé la « Table de correspondance cartouche/ontologie » qui regroupe :

- a) la modélisation telle quelle est implémentée dans l'ontologie du domaine, i.e. pour chaque classe (concept ou relation, i.e. le domaine), son nom, ses attributs et ses rôles le cas échéant ;
- b) les contraintes des rôles, i.e. le « range » ou portée de ce rôle dans l'ontologie, et des attributs, i.e. les types de données comme chaîne de caractère, date, URL, etc. ;
- c) les éléments parmi les classes, rôles et attributs dont les valeurs seront automatiquement renseignées par le moteur d'extraction linguistique ;
- d) le nom de la ou des étiquette(s) sémantique(s) dont la valeur textuelle sera utilisée pour créer les valeurs des éléments concernés, i.e. les nœuds déclencheurs ;
- e) la condition contextuelle implémentée dans la règle d'acquisition ;
- f) les commentaires éventuels provenant des différents échanges entre les intervenants concernés pour garder une trace des problèmes soulevés et des solutions choisies.

Ce document tient une place importante dans la méthodologie OntoPop puisqu'il sert non seulement à la compréhension du fonctionnement du futur couplage, mais également à toute la gestion et maintenance de ce dernier. Il doit obligatoirement être mis à jour à chaque modification de l'ontologie ou de la structure des arbres conceptuels du domaine.

### 6.3.3 *La Phase de Couplage :*

Cette phase consiste à définir et implémenter les Règles d'Acquisition de Connaissance pour l'application documentaire sur la base du modèle de l'ontologie du domaine et de la structure des arbres conceptuels générés par la cartouche linguistique du domaine. Comme nous venons de le voir, chaque élément de l'ontologie utilisé pour l'annotation documentaire ou pour l'enrichissement de la base de connaissance a été répertorié dans la Table de correspondance cartouche/ontologie. En effet, il ne s'agit pas d'instancier ou d'annoter avec tous les éléments définis dans l'ontologie. Certains d'entre eux sont utilisés à d'autres fins par l'application documentaire, comme proposer des fonctionnalités accessibles via l'interface utilisateur de l'application. Outre la Table de correspondance, le responsable de couplage dispose des spécifications détaillées des arbres conceptuels fournies par le linguiste ainsi que la documentation de l'ontologie réalisée par l'ontologue.

Avant d'initier la comparaison des éléments répertoriés de l'ontologie avec les étiquettes linguistiques des arbres, il est nécessaire de constituer un corpus de tests unitaires. Ce corpus, composé d'une sélection de ressources documentaires balayant l'ensemble des arbres conceptuels générés par la cartouche, permet de tester tous les cas de couplage possibles. Le responsable de couplage peut alors analyser en détail chacun des arbres conceptuels générés à partir des tests unitaires. Dans un premier temps, il identifie les étiquettes linguistiques dont les valeurs textuelles correspondent à la valeur d'une instance ou d'une annotation. Ce seront les nœuds déclencheurs d'une nouvelle Règle d'Acquisition de Connaissance. Dans les cas où il y a ambiguïté entre plusieurs étiquettes linguistiques, il analyse le contexte de ces étiquettes qui déterminera les indices contextuels de la règle. Les spécifications détaillées fournies par le linguiste permettent au responsable de couplage d'obtenir rapidement une vue d'ensemble des différents tests unitaires et des étiquettes pouvant jouer un rôle dans la résolution des ambiguïtés.

Dans certains cas, les spécifications détaillées et les tests unitaires ne sont pas suffisants pour formuler les différentes règles d'acquisition. Le responsable de couplage doit également prendre

en compte des règles de gestion métier spécifiées par l'expert. Par exemple, toujours dans le domaine de l'édition juridique (projet Exinis), lors de l'identification d'une décision de jurisprudence, la date de la décision peut apparaître de plusieurs manières dans l'arbre de concept : « /DateDecision », « /DateLecture » ou encore « /DateSeance ».

Or ces trois étiquettes n'ont pas la même importance. Si plusieurs de ces étiquettes sont représentées dans l'arbre de concept d'une décision de jurisprudence, une seule de ces étiquettes possède la bonne valeur de la date de la décision et seul l'expert peut apporter cette connaissance métier : l'étiquette « /DateDecision » prime sur les deux autres, ensuite l'étiquette « /DateLecture » prime sur l'étiquette « /DateSeance », la valeur de cette dernière n'est donc exploitée que si aucune autre des étiquettes représentant une date n'est présente dans l'arbre conceptuel d'une décision de jurisprudence. Trois règles d'acquisition traduisent cette règle métier, dans lesquelles la primauté d'une étiquette sur une autre est déterminée par le niveau de confiance accordée à chacune des règles.

D'autres règles de gestion métier peuvent porter sur la normalisation de la valeur générée en sortie. Celle-ci peut être une concaténation de différentes valeurs d'étiquettes ou une combinaison de noms d'étiquettes et de leurs valeurs ou encore une constante, etc.

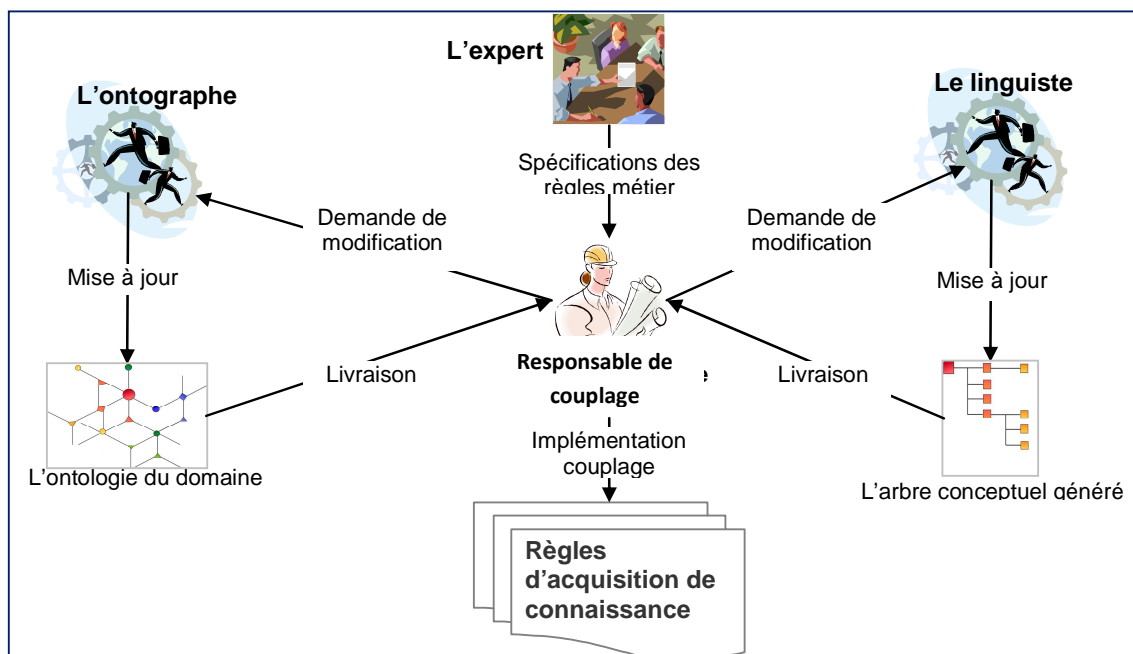


Figure 20 : Analyse des problèmes rencontrés par le responsable de couplage (d'après Amardeilh, 2007).

Une fois que le responsable de couplage a spécifié dans sa Table de correspondance l'ensemble des nouvelles règles d'acquisition, il peut les implémenter via l'Editeur de Règle. Pour chacune d'entre elles, il entre les informations requises à sa création. Les règles peuvent alors être testées une par une ou toutes à la fois. La première option permet de mieux contrôler que les résultats produits par chacune des règles correspondent bien aux attendus. La seconde option permet de détecter des conflits entre plusieurs règles ainsi que les incohérences. Si les règles ne produisent pas les résultats attendus, divers facteurs peuvent en être la cause : une mauvaise configuration de la règle dans l'Editeur, des incohérences dans les arbres conceptuels, des problèmes de formats avec le modèle de l'ontologie, des oublis, des conflits, etc. Le responsable de couplage doit être capable d'analyser l'origine du problème pour savoir s'il peut y remédier en corrigeant la règle déviante ou bien si le problème nécessite une modification de l'ontologie ou de la cartouche linguistique (Figure 20, page précédente).

Auquel cas, il demande à l'ontologue ou au linguiste d'effectuer les modifications nécessaires dans leurs modèles. Ces modèles ne sont donc pas fixés une fois pour toute à ce stade de la méthodologie, mais plutôt constamment retravaillés jusqu'à la validation définitive effectuée à l'étape suivante. A chaque nouvelle règle d'acquisition testée et approuvée sur le corpus de tests unitaires, le responsable de couplage met à jour la Table de correspondance cartouche/ontologie. A l'issue de cette phase, cette Table de correspondance est de nouveau livrée à l'ensemble des partenaires, afin qu'ils puissent prendre connaissance des Règles d'Acquisition de Connaissance ayant été définies par le responsable de couplage. La prochaine étape concerne la validation des résultats produits par l'ensemble de ces règles d'acquisition.

#### *6.3.4 La Phase de Validation :*

Cette phase consiste à mesurer la qualité de l'ensemble de l'application, et plus particulièrement du couplage, par rapport aux objectifs fixés initialement (Figure 21, page suivante). Pour cela, la cartouche linguistique du domaine, puis le couplage de cette cartouche avec l'ontologie du domaine doivent être testés afin de contrôler la capacité d'intégration des modèles produits et la pertinence des solutions proposées. A cet effet, un corpus de validation est constitué à partir des documents non utilisés dans le corpus de tests unitaires de l'étape précédente. Ce corpus de validation est utilisé pour les deux processus de validation.

Le premier mesure la performance de l'outil linguistique, la qualité de ses extractions, sa couverture vis-à-vis du domaine concerné, etc. Pour ce faire, le linguiste calcule la précision, c'est-à-dire le silence (éléments oubliés), et le rappel, c'est-à-dire le bruit (éléments pris en compte par erreur), généré par les patrons d'extraction sur les documents du corpus. Il compare ces résultats aux tests unitaires réalisés précédemment et contrôle que les extractions produites sont conformes à celles attendues. Les taux de précision et de rappel doivent atteindre un seuil suffisamment important pour être acceptables par l'application. Ce seuil est défini conjointement par le linguiste et l'expert au vu de la complexité du domaine étudié et des besoins de l'application documentaire.

Le second processus de validation analyse la qualité des propositions fournies par les Règles d'Acquisition de Connaissance, tant au niveau du peuplement de l'ontologie que de celui de l'annotation sémantique. Ce processus de validation utilise le Module d'Annotation et d'Acquisition et accessoirement l'interface de validation, qui permet de visionner les résultats produits tant au niveau du peuplement d'ontologie que de l'annotation sémantique pour tout document traité.

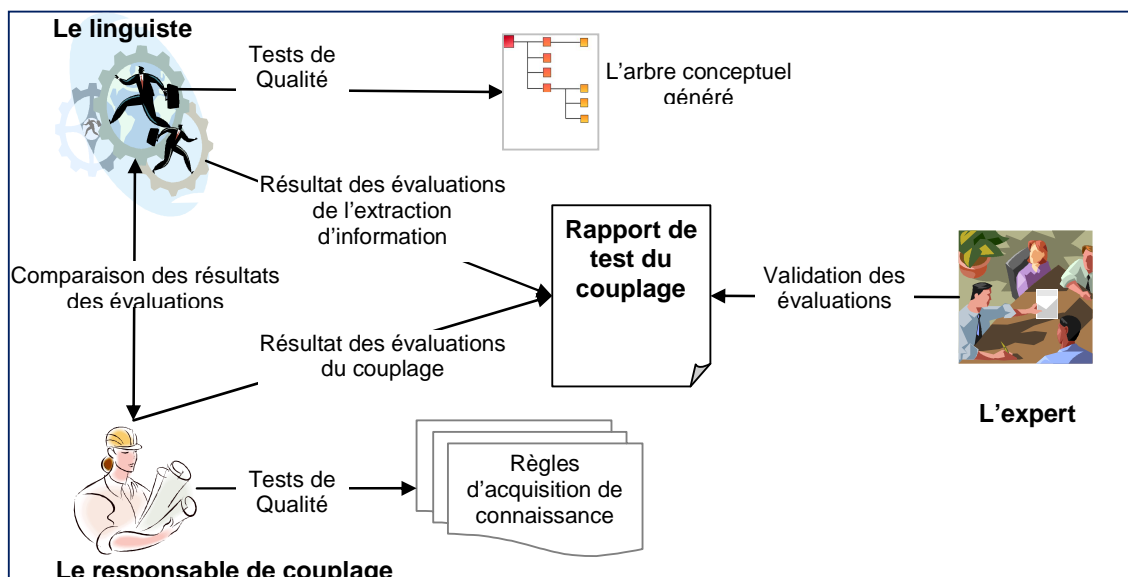


Figure 21 : Phase de validation (d'après Amardeilh, 2007).

La comparaison des résultats des deux processus de validation a lieu à plusieurs reprises, lors de réunions de travail entre le linguiste et le responsable de couplage. Si les résultats sont jugés insuffisants vis-à-vis des objectifs fixés dans les spécifications détaillées de l'application, l'origine de la défaillance doit être réparée : les règles d'acquisition, la cartouche linguistique ou l'ontologie

du domaine. Dans le premier cas, le responsable de couplage modifie les règles afin de corriger les erreurs de conflits, d'oublis, de syntaxe, etc. et recommence son évaluation à partir des cas posant problème. Dans le second cas, il travaille avec le linguiste pour ajuster les patrons d'extraction ambigus. Dans le dernier cas, il demande à l'ontologue de procéder à des modifications dans la modélisation de l'ontologie du domaine. Si les modifications demandées, tant au linguiste qu'à l'ontologue, sont mineures, de nouvelles versions de la cartouche ou de l'ontologie sont livrées à le responsable de couplage pour réitérer cette phase de Validation. Si ces modifications sont majeures alors le responsable de couplage doit reprendre à partir de la phase de Couplage, voire à partir de la phase de Structuration avec validation des modifications par l'expert. Par contre, si les résultats sont jugés satisfaisants, le rapport de test du couplage et du taux de couverture entre les outils est remis à l'expert au cours d'une réunion où une démonstration de l'application est effectuée. Si l'expert valide ces résultats, alors l'application est livrée, installée et mise en service à l'étape suivante.

### 6.3.5 La Phase de Mise en Service :

Cette phase consiste à livrer, installer et assurer le suivi des différents composants de l'application documentaire (figure 22) :

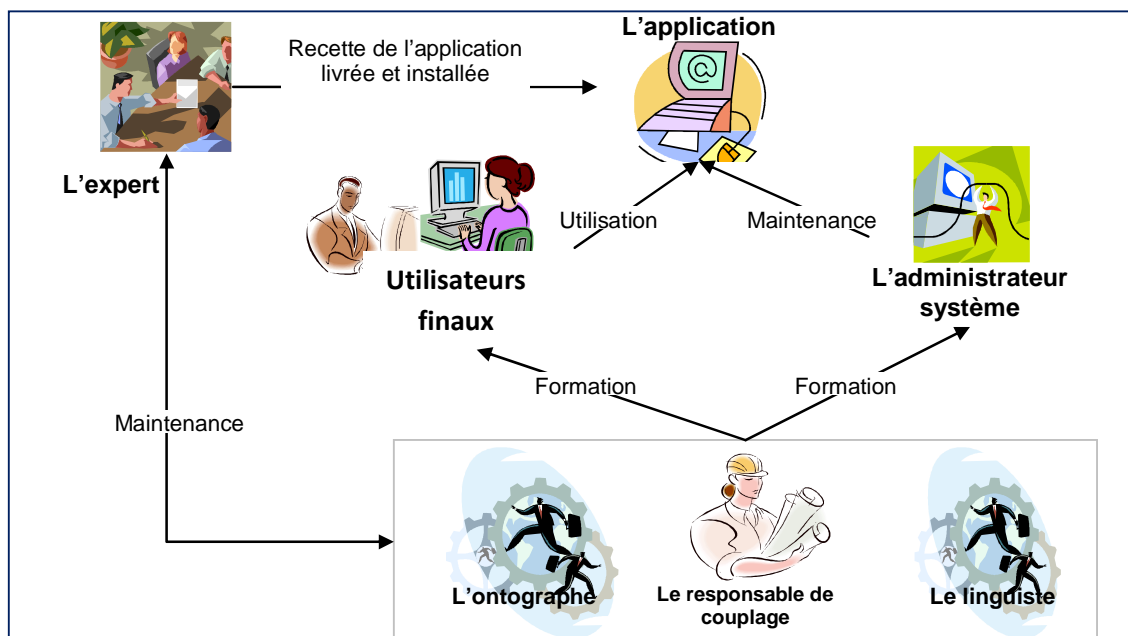


Figure 22 : Phase de mise en service (d'après Amardeilh, 2007).

Une fois que toute l'installation est prête, l'expert procède à la recette de l'application en vérifiant chacune des fonctionnalités demandées dans les spécifications détaillées de l'application ainsi que la cohérence et la qualité des résultats produits, comparativement à ceux présentés à l'étape précédente. Si l'installation de l'application montre des défauts, les corrections appropriées sont apportées par les divers intervenants dans les plus brefs délais. Dans le cas contraire, l'application est mise en service auprès des utilisateurs finaux après qu'ils aient suivi une à plusieurs formations selon le degré de changement apporté par la nouvelle application. Ces formations concernent notamment la navigation, la recherche et la publication dans le référentiel de l'application ainsi que la validation des résultats fournis par l'application dans le cas d'un processus semi-automatisé. Outre les utilisateurs finaux, une formation est également dispensée au futur administrateur système de la nouvelle application. Ce dernier doit apprendre à installer, gérer et maintenir chaque outil et chaque composant formant l'ensemble de l'application. Enfin, le linguiste, l'ontologue et le responsable de couplage peuvent proposer à l'expert un ensemble d'améliorations possibles de l'application sur la base de leurs analyses et de leurs travaux développés tout au long de la méthodologie OntoPop. Par exemple, le linguiste aura peut-être identifié de nouveaux concepts à exploiter grâce à la lecture attentive des documents du corpus représentatif du domaine. A la vue de ces nouvelles étiquettes sémantiques, le responsable de couplage peut alors demander à l'ontologue la manière de modéliser ces nouveaux concepts dans l'ontologie du domaine. Cette nouvelle modélisation est alors proposée à l'expert pour savoir si le peuplement de ces nouveaux concepts ou l'annotation des documents par ces nouveaux concepts peuvent être considérés comme pertinents pour les utilisateurs finaux. Si ces ajouts sont approuvés par l'expert alors ils feront soit l'objet d'une maintenance dans le cas de modifications mineures, soit l'objet d'une nouvelle version de l'application dans le cas contraire.

#### **6.4 Conclusion à propos de la méthodologie OntoPop :**

La méthodologie OntoPop a été définie pour résoudre des problèmes rencontrés par les acteurs du projet PressPro. Elle décrit le déroulement idéal d'un projet couplant moteur d'extraction d'information et système de représentation des connaissances via la passerelle OntoPop. La description de cette méthodologie nous permet de mieux saisir l'interdépendance des tâches, et l'ampleur des interactions nécessaires à la conception d'une application reposant sur un tel couplage. Au travers de l'importance du rôle attribué à l'« expert », cette description met aussi en évidence le degré d'« adhérence » de l'information détenue par le client, information qu'il est nécessaire de transférer à Temis et à Mondeca, pour la conception de l'application recherchée.

Nous revenons maintenant au projet Exinis, projet au cours duquel cette méthodologie a contribué à structurer le travail des différents acteurs.

## **7 LE PROJET EXINIS :**

Alors que le projet PressPro était en cours de développement, un nouveau projet de création automatisée de base de connaissance dans le secteur de la presse et de l'édition a été lancé, chez Exinis. Nous allons maintenant décrire ce projet, que nous appellerons « projet Exinis », qui a, de nouveau, associé Temis et Mondeca. Comme pour le projet PressPro, nous aborderons successivement l'origine du projet, la description du projet et sa chronologie, les phases amont du projet jusqu'à la signature du contrat, l'organisation du projet, les phases de réalisation de celui-ci, et en fin nous en présenterons une synthèse.

### **7.1 L'origine du projet Exinis :**

Ce projet a été initié à la demande d'un groupe français, spécialisé dans l'édition juridique, filiale d'un groupe international, Exinis Group, qui publie une gamme de produits d'information juridique de référence (encyclopédies, revues et CD-ROM). Exinis France met notamment à disposition des professionnels du droit toutes les décisions de jurisprudence depuis 1960. Exinis France doit donc traiter un gros volume d'informations provenant de sources extérieures à l'entreprise, informations dont le volume ne cesse d'augmenter. Le projet que nous avons suivi, s'est inscrit dans le cadre du projet « Kepler », qui visait à la refonte complète du système éditorial d'Exinis Group. Fondamentalement, l'objectif du projet Kepler était de passer de la « gestion de contenu » à la « gestion de la connaissance (que recèle le contenu) ». Le projet Kepler était donc un projet de grande ampleur qui nécessitait le développement d'un framework, i.e. d'une plateforme informatique fondée sur différentes briques logicielles indépendantes et réutilisables.

A l'origine du projet, on trouve JR, informaticien et « architecte innovation » de la solution, qui avait été recruté, en 2000, par Exinis France pour transférer sur Internet les contenus qu'il mettait à disposition de ses clients. Par ailleurs, Exinis, qui cherche en permanence à accélérer et à automatiser ses processus d'acquisition d'information, disposait d'une base de données, ou « base de liens », appelée Sara. Cette base était alimentée par une application permettant d'extraire des chaînes de caractères. Sara permettait d'indexer les documents juridiques, et de placer des liens hypertextes reliant les documents juridiques entre eux. Mais, en 2004, l'architecte innovation, qui

avait auparavant travaillé avec des chercheurs en sciences du langage, et connaissait les perspectives offertes par les systèmes de représentation des connaissances, a eu l'idée de remplacer ce système par une base de connaissance, elle-même alimentée par une application d'extraction d'information. Le projet que nous avons étudié concernait Exinis France, mais il était susceptible de trouver des prolongements dans les différentes filiales internationales d'Exinis Group.

## 7.2 Description et chronologie du projet :

Quand l'architecte innovation d'Exinis s'est intéressé à la question de la gestion des connaissances, il a cherché un outil permettant de représenter sous forme de base de connaissance, l'information juridique utile au métier de rédacteur juridique. Il a ainsi identifié l'éditeur de logiciel Mondeca, et découvert qu'il proposait une solution couplée avec la solution d'extraction d'information proposée par Temis. Depuis son arrivée en 2000, JR estimait qu'Exinis devait chercher à utiliser les ressources offertes par le traitement du langage, et ne pas se limiter au traitement de chaînes de caractères. Jusqu'alors, les informaticiens d'Exinis travaillaient en effet de manière récurrente à l'optimisation du système de gestion de base de liens (liens entre un « contenu », i.e. un texte, et les références qu'il cite), une base de données et non une base de connaissance, via l'optimisation du traitement des chaînes de caractères. Ainsi, le projet imaginé par JR, s'inscrivait bien dans le cadre du projet Kepler. Il nécessitait la mise en œuvre chez Exinis de deux sauts technologiques :

- La prise en compte du langage au lieu des chaînes de caractères,
- Le passage de la gestion des données à la gestion des connaissances.

Il s'agissait alors de passer « à un autre mode de pensée, un autre mode de spécification » (JR, entretien juin 2006). Cela impliquait un certain niveau de compréhension du métier de rédacteur juridique, de la part des acteurs en charge du développement de la solution.

Le projet Exinis, inscrit dans le cadre du projet Kepler, visait fondamentalement à fournir à ses rédacteurs, une unique source d'information, le référentiel ou système de représentation des connaissances ITM de Mondeca. Le projet a ainsi été mis en place avec les objectifs suivants :

**1. Balisage des nouveaux articles :** chaque rédacteur juridique qui écrit un nouvel article fait un certain nombre de références à des textes de lois, décisions de jurisprudence, et autres publications officielles. Ces références doivent être balisées et annotées dans l'article en fonction de leurs propriétés, comme le nom du code ou du texte de loi, sa date de publication, son émetteur, etc. La solution à développer devait automatiser un tel balisage dans chaque article en cours de rédaction. Cela passait par le remplacement de l'ancien système, Sara, en place chez Exinis France depuis trente ans. La « base de liens » Sara était une base de données qui rassemblait tous les liens entre la jurisprudence et la législation. Le projet Exinis visait donc à faire migrer cette base, comprenant environ quinze millions de liens, vers la base de connaissance ITM. Une fois le nouveau système en production, les briques logicielles fournies par Temis devaient extraire la connaissance des documents en cours de rédaction, grâce au logiciel IDE couplé à une « cartouche juridique », permettant ensuite le balisage automatique des documents, après validation automatique par l'ontologie intégrée à la base ITM. En cas de doute, un écran de validation devait apparaître, permettant ainsi au rédacteur d'intervenir. Des liens entre contenus devaient ainsi être créés. Une fois le texte balisé, il devait être possible de créer une publication avec des liens hypertextes, autorisant ainsi l'accès direct aux différentes références citées dans cette publication.

**2. Acquisition des nouvelles décisions de jurisprudence:** le deuxième volet du projet Exinis, concernait l'acquisition des nouvelles décisions de jurisprudence. Cette acquisition devait être elle-aussi optimisée, grâce à l'extraction automatisée de chaque propriété identificatoire de ces décisions comme la juridiction, le siège, la formation, la date de la décision, la décision elle-même, les parties concernées, etc. La solution devait donc extraire et transférer dans la base de connaissance ces propriétés identificatoires. Cela nécessitait le développement d'une autre cartouche de connaissance.

**3. Optimisation de la recherche sur le portail Internet d'Exinis :** il devait y avoir ultérieurement un autre volet. Il concernait la recherche d'information sur le site web d'Exinis. La recherche d'information sur ce site pouvait être elle-aussi optimisée par le recours à une base de connaissance, alimentée par un moteur d'extraction d'information.

Idéalement, tous les rédacteurs, « éditeurs » ou « auteurs », travaillant pour Exinis France devaient utiliser à terme le nouveau système. On comptait environ six mille auteurs (dont cinq cents réellement actifs), et quatre cents éditeurs, ces derniers chargés d'agréger les différents contenus

et responsables de la validité des contributions fournies par les auteurs. Mais, l'ensemble des personnes censé utiliser, à court terme, le nouveau système était évalué à une cinquantaine, par Exinis.

Comme PressPro, Exinis a fait appel aux éditeurs de logiciel Temis et Mondeca. Exinis a donc réalisé un investissement lourd pour faire développer une première « cartouche de connaissance juridique » (ou « cartouche juridique ») par Temis, pour faire réaliser l'ontologie et rédiger des règles d'acquisition de connaissance personnalisées par Mondeca. Il a aussi fait appel à une société de services 4D Concept, pour faire développer le « système éditorial » (que nous appellerons « application documentaire », comme dans le projet PressPro), i.e. l'application informatique qui constitue l'interface utilisateur et fédère les différentes briques logicielles.

### **Chronologie du projet :**

Le projet Exinis s'est déroulé selon un processus moins exploratoire et chaotique que le projet PressPro, et a suivi un déroulement dont les étapes peuvent être identifiées plus distinctement. Ces étapes correspondent à peu de choses près à ce qui était prévu initialement, quant à la nature et à la succession des étapes, mais pas en ce qui concerne les délais. La mise en œuvre de la méthodologie « OntoPop » a été effective à partir de juillet 2006, et du travail sur la version V1 :

1. Premier contact en avril 2004.
2. Signature d'un premier contrat en juin 2004 pour un « démonstrateur ».
3. Livraison du « démonstrateur » en octobre 2004.
4. Contrat définitif signé pour la solution complète, le 10 mai 2005.
5. Version V0 (pré production) livrée en février 2006 (prévue pour octobre 2005) incluant les logiciels IDE et IDK de Temis, une « cartouche juridique » spécialement développée par Temis, le logiciel ITM et le développement de RAC spécifiques par Mondeca.
6. Projet mis en stand by de janvier à juin 2006.
7. Relance du projet en juin 2006 et mise en œuvre de la méthodologie OntoPop,
8. Version V1 de la cartouche de connaissance (production), livrée en février 2007 (initialement prévue pour fin 2005).

9. Version V1 de la base de connaissance par Mondeca, alimentée par la base de liens Sara, livrée en octobre 2007
10. Version V1 du système éditorial livré par 4D Concept et Mise en pré-production du système en décembre 2007.
11. Mise en production du système en juin 2008.

### **7.3 Les phases amont jusqu'à la signature du contrat définitif :**

#### *7.3.1 Premiers contacts en avril 2004 :*

L'origine du projet Exinis remonte à avril 2004, moment où le CIO (Chief Information Officer) d'Exinis France, a donné son feu vert à JR, informaticien et futur « architecte innovation » du projet, pour réaliser, en six mois, un « démonstrateur », i.e. un logiciel susceptible de démontrer l'intérêt du système imaginé. En avril 2004, le choix de Mondeca par Exinis a résulté du fait qu'il fallait que JR montre rapidement un outil qui alimente une base de connaissance, or celui-ci avait déjà travaillé avec Mondeca, à propos de la gestion de référentiel métier. Le choix de Mondeca par JR alors qu'il existait d'autres logiciels de gestion de base connaissance, dans des laboratoires de recherche ou chez un éditeur de logiciel Anglais, résulte aussi de ce que Mondeca avait « le même langage, la même vision globale de la gestion des connaissances » (entretien avec JR, juin 2006) que lui. Après avoir retenu Mondeca, Temis devenait incontournable comme éditeur de logiciel spécialisé dans le *text mining*, puisque que déjà partenaire de Mondeca sur des solutions de *publishing* (PressPro). Il en est allé de même pour 4D Concept, pour ce qui est du rôle d'intégrateur de la solution car cette petite société de services avait déjà travaillé pour Exinis, avec l'architecte innovation.

#### *7.3.2 Un premier contrat pour un démonstrateur en juin 2004 :*

Cette première étape est marquée par la volonté de l'architecte innovation d'Exinis, ainsi que des chefs de projet de Temis et de Mondeca, de procéder à une intégration rapide des ébauches de briques logicielles constitutives de la future solution, afin de faire faire des tests et d'obtenir un retour d'information lui aussi rapide de la part du client. Un premier contrat a donc été signé en juin 2004, entre Exinis France d'une part et quatre « prestataires » (terminologie employée par Exinis) d'autre part pour le développement d'un « démonstrateur » (logiciel montrant de manière

simplifiée le fonctionnement du futur logiciel). Les quatre prestataires étaient Temis, Mondeca, 4D Concept, mais aussi Acamaya pour le logiciel e-room (outil de travail collaboratif finalement inutilisé et rejeté, mais que la direction éditoriale d'Exinis voulait au départ absolument intégrer). Il fallait montrer à la direction d'Exinis Group ce que l'innovation pourrait apporter.

Le démonstrateur a été développé en moins de six mois et terminé mi-octobre 2004. En cette occasion, les acteurs du projet chez les « prestataires » ont travaillé avec, du côté d'Exinis France, JR et deux personnes à la fois utilisateurs (rédacteurs juridiques) et dotées de compétences en représentation des connaissances, dont la connaissance des usages était indispensable. JR a dirigé le projet « démonstrateur », mais les informaticiens d'Exinis n'ont pas participé à son développement.

### 7.3.3 Présentation du démonstrateur à Exinis US :

Une présentation du démonstrateur a été faite mi-octobre 2004, par JR et deux personnes d'Exinis France à leurs collègues d'Exinis US (alter egos de JR et personnes hiérarchiquement supérieures, du secteur éditorial et de l'informatique). Pour JR, le pari était difficile car Exinis Group avait dépensé quarante-cinq millions de dollars, sur un projet que l'innovation proposée allait rendre obsolète. Le démonstrateur fonctionnait réellement (extraction, puis transfert vers une base de connaissance), même si la partie « couplage » était encore limitée. Cette présentation a été déterminante car les responsables d'Exinis US ont alors compris le potentiel d'une telle solution, « cela les a fait rêver » (JR, entretien 2006). Dans le cadre du projet Kepler, Exinis France est alors devenu leader dans le domaine du *text mining*, et de la gestion des connaissances. Il en a aussi découlé différents projets impliquant certaines filiales d'Exinis Europe (Exinis D et Exinis UK), le projet « Kepler » s'est ainsi étoffé.

Les enseignements retirés du développement du démonstrateur pour Exinis furent principalement les suivants :

1. Le démonstrateur confirmait les hypothèses de l'architecte innovation d'Exinis,
2. Il a mis en évidence la complexité potentielle du couplage entre l'extraction d'information (Temis) et la gestion des connaissances (Mondeca),
3. Il a mis en évidence les performances et limites du système : il a permis de comprendre que la base de connaissance à construire allait être extraordinairement volumineuse. JR a

alors pensé que la difficulté principale viendrait, non des données à traiter, mais de ce que la technologie employée par Mondeca n'avait pas encore atteint une maturité suffisante concernant le stockage d'une base de connaissance en base de données relationnelle, notamment du point de vue de la gestion des requêtes que cela implique. JR estimait que cette question faisait de ce projet, pour la partie Mondeca, un projet de recherche plus que de développement.

#### 7.3.4 *Le contrat définitif signé en mai 2005 :*

Le contrat définitif, relatif au projet Exinis, a été signé officiellement en mai 2005, entre Temis/Mondeca/4D Concept/Unilog et Exinis, par le PDG d'Exinis (France). Mais les deux éditeurs de logiciel, Temis et Mondeca, y travaillaient déjà depuis mars 2005. Au moment du démarrage du projet, Mondeca étant absorbé par la fin du projet PressPro, c'est en conséquence Temis qui fut promu leader du projet Exinis. L'autre élément qui a conduit à la mise en avant de Temis sur ce projet, est que Temis était une société plus importante que Mondeca, qui, de plus, bénéficiait d'une certaine aura due à son partenariat avec Xerox : tout cela a contribué à rassurer la direction d'Exinis. Pour ce projet, le nombre de jours homme dédiés devait être de 140 pour Temis, 150 pour Mondeca et environ 200 pour 4D Concept.

En ce qui concerne les spécifications du projet, Exinis, en l'occurrence JR, estimait qu'elles possédaient un niveau minimum de précision et qu'il s'agissait de spécifications « gros grain » pour Exinis. A l'inverse, Temis et Mondeca les ont jugées relativement précises, surtout en comparaison de celles qui avaient été rédigées pour PressPro. Le comité de projet d'Exinis France estimait savoir exactement le niveau de qualité d'extraction qu'il attendait pour ce projet, mais les taux d'extraction attendus n'ont pas été définis précisément. Tous les acteurs s'accordaient toutefois sur le fait que ces spécifications permettaient d'avoir une vision relativement claire du « périmètre » de la solution.

#### **Les tâches à accomplir par Temis, Mondeca et 4D Concept :**

- Temis devait définir une première cartouche de connaissance juridique pour le « balisage de références »,
- En parallèle, Mondeca devait définir l'ontologie du domaine, et les RAC correspondantes,

- Ensuite, ces deux premières tâches étant réalisées, la cartouche de connaissance devait permettre de transférer la base de liens utilisée par Exinis dans la base ITM,
- En parallèle de ces différentes tâches, 4D Concept devait développer le système éditorial (ou application documentaire), un environnement de rédaction, permettant aux rédacteurs d'Exinis de rédiger puis de baliser leurs articles.

La direction d'Exinis France a estimé qu'un autre intégrateur informatique devait travailler à l'intégration du projet, en plus de 4D Concept. Il s'agissait de les faire travailler ensemble, voire de les mettre en concurrence. Exinis France n'a alors identifié que deux prestataires possédant des compétences relatives aux technologies employées par Temis et Mondeca, Unilog et Sword. Sword n'ayant pas su répondre correctement à la sollicitation d'Exinis, il ne restait alors qu'Unilog. Cela a repoussé le début officiel du projet au mois d'avril 2005, alors que Temis avait déjà commencé à travailler sur sa cartouche de connaissance juridique. Ces sociétés de services informatiques présentaient des caractéristiques propres et ont dû assumer des tâches distinctes :

- 4D Concept, SSII spécialisée et de petite taille, qui connaissait les usages des rédacteurs chez Exinis, devait développer l'application documentaire, et à ce titre, entre autres, réaliser les interfaces de recherche et de validation destinées aux utilisateurs finaux ;
- Unilog, société de services de grande taille, ne possédant pas la même connaissance des usages que 4D Concept, qui a été chargé de la « conduite du changement », i.e. concrètement d'organiser les ateliers où les utilisateurs devaient formuler des avis relatifs à des choix de conception.

#### **7.4 Les acteurs et l'organisation du projet (à partir de mai 2005) :**

Nous décrivons ici les rôles des principaux acteurs du projet, plaçant entre guillemet les terminologies employées par les acteurs eux-mêmes. Ensuite, nous présentons les différents types de réunion qui ont structuré le projet.

##### *7.4.1 Les acteurs du projet chez Exinis :*

Un certain nombre de rôles ont été attribués aux différents acteurs du projet d'Exinis France, que nous présentons ci-après :

- **Jean-Pierre**, le Chief Information Officer (CIO) d'Exinis France : responsable de l'ensemble des décisions majeures du projet, il y a consacré environ 10% de son temps.

- **Alexandre**, le « sponsor du projet », directeur des Equipes Editoriales : informé de l'avancement du projet, il pouvait être amené à gérer, avec le comité de direction, des décisions qui dépasseraient le strict cadre du projet. Le sponsor était aussi censé transmettre la « bonne parole » relative au projet, dans l'entreprise.

Viennent ensuite, les acteurs, moins haut placés dans la hiérarchie de l'entreprise Exinis France, qui ont été les plus impliqués dans le projet :

- **JR**, l'« architecte innovation » et « responsable informatique » : informaticien à l'origine du projet, il a eu la charge de le superviser et d'en fixer les orientations ; il a été à ce titre l'interlocuteur privilégié des chefs de projet des « prestataires » (Temis, Mondeca et 4D Concept). Il a dirigé une équipe informatique de cinq personnes, pratiquement dédiée au projet, dont Vassili en charge des questions d'installation et de serveur, et Tim en charge de la qualité du système, Christophe en charge de la maintenance des RAC, après que celles-ci ont été définies.

- **Olivier**, le « chef de projet intégration informatique » : informaticien spécialiste des bases de données, qui devait s'occuper du suivi et de la gestion opérationnelle du projet ; il a surtout travaillé avec l'« architecte innovation » du projet, la « responsable éditoriale » et « rédactrice experte » Laurence, et l'expert ontologue de Mondeca, Bernard, sur la définition de l'ontologie.

- **Laurence et Emilie**, deux « utilisateurs experts » ou « rédactrices expertes » : toutes deux rédactrices juridiques, membres de l'équipe éditoriale, dotée d'une formation en gestion documentaire, et membres du « comité de projet ». Emilie a collaboré avec la responsable de couplage de Mondeca à la définition des RAC. Laurence était la supérieure hiérarchique d'Emilie.

- Un « **comité de projet** » composé de **cinq personnes** (les deux rédactrices expertes Laurence et Emilie, l'« architecte innovation », le « chef de projet », et Tim en charge de la qualité du système).

- Une « **équipe éditoriale** » : une **quinzaine de rédacteurs juridiques** (éditeurs et quelques auteurs), utilisateurs du système participant régulièrement à des « ateliers utilisateurs ».

#### 7.4.2 Les acteurs du projet chez Temis :

Nous retrouvons ici à peu près les mêmes acteurs que pour le projet PressPro :

- **Sylvie**, Temis Paris, linguiste et informaticienne, « chef de projet Temis », dans un premier temps.
- **Jean-Pierre**, Temis Paris, linguiste, a assumé le rôle de chef de projet cartouche de connaissance, et expert linguiste ; il a aussi, dans un deuxième temps, succédé à SG comme chef de projet Temis.
- **Amandine**, Temis Paris, linguiste et informaticienne, a travaillé au développement informatique de la cartouche de connaissance. AM lui a succédé après son départ de Temis en 2007.
- **Christian**, Temis Paris, informaticien, a travaillé sur la partie catégorisation/indexation du logiciel, il a assumé le rôle de chef de projet catégorisation.

Comme pour le projet PressPro, d'autres ont joué un rôle plus ponctuel, ne consacrant qu'une faible partie de leur temps au projet :

- **Vincent**, Temis Grenoble, ingénieur avant-vente devenu « *account manager* » pour les clients de Temis situés dans le secteur Presse/Edition (appelé « *publishing* ») : il a joué un rôle d'interface commerciale avec Exinis, lors des négociations qui ont jalonné le projet.
- **Karene**, Temis Grenoble, linguiste et informaticienne, qui a travaillé au développement du « démonstrateur ».
- **Olivier**, Temis Grenoble, Core Product Technical Manager, responsable du suivi et de l'évolution des logiciels de Temis. Il a participé à la rédaction des spécifications.

### 7.4.3 Les acteurs du projet chez Mondeca :

De manière similaire au projet PressPro, les acteurs du projet Exinis chez Mondeca ont été :

- **Bernard**, expert ontologue ou « ontologue », appelé « consultant » en interne, spécialiste de la gestion des connaissances et responsable de la définition de l'ontologie, a assumé le rôle de chef de projet ontologie et d'expert ontologue.
- **Benoit**, chef de projet Mondeca, informaticien formé à la gestion des connaissances,
- **Florence Amardeilh**, informaticienne elle aussi formée à la gestion des connaissances, chef de projet couplage ITE/ITM et chargée de la définition des RAC,
- **Thomas**, informaticien et « chef de produit [technique] ITM »,
- **Anh**, informaticien, tous deux en charge du développement des modules logiciels de la passerelle OntoPop.

Comme pour le projet PressPro, d'autres ont joué un rôle plus ponctuel, ne consacrant qu'une faible partie de leur temps au projet :

- **Gilles**, responsable commercial, qui a assuré la coordination avec le client et Vincent de Temis pour les phases contractuelles,
- **Jean Delahousse**, président, qui supervisait l'avancement du projet.

### 7.4.4 Les acteurs du projet chez 4D Concept :

Un consultant, chef de projet du « système éditorial » (ou application documentaire), a travaillé avec deux informaticiens de 4D Concept au développement de ce système.

### 7.4.5 L'acteur du projet chez Unilog :

Un consultant, chef de projet « conduite du changement » était initialement en charge de l'organisation des ateliers utilisateurs. Il a assez rapidement quitté le projet, Unilog ne faisant dès lors plus partie du projet dès l'automne 2005.

#### 7.4.6 Les réunions de projet organisées par l'architecte innovation d'Exinis :

Les contacts entre acteurs du projet furent nombreux et permanents pendant toute la durée du projet. Cela n'empêcha pas la tenue d'un grand nombre de réunions de projet. Ces réunions étaient soit planifiées, soit organisées en fonction des problèmes rencontrés sur les projets. Nous listons ici les réunions organisées chez Exinis France sur une base régulière :

1. Réunions hebdomadaires d'un comité de projet très réduit : JR, architecte innovation, Olivier, chef de projet, et Emilie, rédactrice experte ;
2. Réunions hebdomadaires de ce comité de projet réduit avec 4D Concept.
3. Réunions mensuelles de « pilotage », rassemblant le « comité de projet » au complet et les chefs de projet des « prestataires » (Mondeca, Temis et 4D Concept).
4. Les « ateliers utilisateurs »: ils réunissaient le « comité de projet » et « l'équipe éditoriale ». Le « comité de projet » préparait, quasiment pour chaque atelier, trois alternatives possibles relatives à une question ou un problème particulier, en expliquant aux participants de l'atelier, membres de l'« équipe éditoriale », que les trois options présentées étaient incontournables. Le « comité de projet » essayait ainsi d'amener les participants vers l'alternative qui lui convenait. Ces ateliers avaient été mis en place parce que JR voulait montrer vers quoi le projet allait, et aller aussi vite que possible : il fallait donc, selon lui être directif, sans donner l'impression que les choses étaient imposées. Pendant quatre mois, à partir d'avril 2005, le « comité de projet » a mené plusieurs types d'ateliers, au rythme d'une réunion par mois. Ils ont porté sur les différents aspects de la solution à développer, toutefois il y a eu peu d'ateliers utilisateurs relatifs à la cartouche de connaissance juridique, le sujet s'y prêtant moins que les autres (interfaces utilisateur, ontologie) de par son caractère technique.

### 7.5 Les phases de réalisation du projet (à partir de mai 2005) :

#### 7.5.1 Développement de la version V0 de la solution (mai 2005-février 2006):

##### **La cartouche de connaissance :**

Pour l'équipe-projet de Temis, il s'agissait d'extraire les passages de texte où sont cités des articles de loi. Il s'agissait donc d'une tâche qui pouvait être réalisée par Temis en collaboration avec

Exinis, une tâche en grande partie « découplable » des autres. Ce composant a été développé une première fois en juillet 2005, puis une deuxième version a été fournie en août 2005. Temis a travaillé à l'amélioration de cette version à partir des remarques d'Exinis, jusqu'en février 2006. Cette cartouche était attendue avec méfiance par Exinis, car il s'agissait de la partie la plus « nouvelle » du futur système, mais les premiers livrables ont été accueillis favorablement à l'été 2005. Le développement de la cartouche de connaissance n'a pas été traité au travers des ateliers utilisateurs, car l'architecte innovation d'Exinis estimait qu'elle présentait un caractère trop technique pour les utilisateurs « classiques », et il pensait de plus savoir parfaitement ce qu'il fallait extraire. En plus de l'architecte innovation, qui en orientait le développement, quatre acteurs ont travaillé sur la cartouche de connaissance, les deux rédactrices expertes d'Exinis, et le chef de projet Temis, et Amandine en charge du développement informatique. Ces quatre acteurs ont été qualifiés de « champions » par l'architecte innovation d'Exinis. A l'inverse du projet PressPro, cette tâche a donc été réalisée en collaboration directe et régulière avec deux futures utilisatrices du système.

### **L'ontologie et les RAC :**

Le travail à réaliser par Mondeca s'est avéré plus long et plus compliqué que prévu. Mais il a permis de traiter un certain nombre de problèmes nouveaux, ce qui a par la suite conduit à faire évoluer certains composants logiciels de la passerelle OntoPop, de la base de connaissance ITM, mais aussi à enrichir le concept de RAC tel qu'il avait été défini pour le projet PressPro. Les apports sont donc très importants, et résultent en grande partie du travail effectué par le client, et de l'interaction de celui-ci avec des membres de l'équipe-projet de Mondeca, comme nous allons le décrire maintenant.

Tout d'abord, dès le début du projet, l'architecte innovation a fait développer chez Exinis, une base de connaissance de test qui a rapidement mis en évidence des problèmes de stabilité de cette base, lorsque le volume de connaissance à traiter devenait très élevé, problème dit de « scalabilité ». Il a mis aussi en évidence des problèmes liés à la lenteur des requêtes adressées à cette base de connaissance. De même un couplage de test a mis en évidence la lenteur du couplage entre IDE et ITM, dès lors qu'un flux régulier d'information était à traiter. Dans les deux premiers cas, les problèmes concernaient les composants logiciels de la base ITM, et leur résolution a nécessité une collaboration poussée entre le chef de produit ITM de Mondeca, et le chef de projet d'Exinis, spécialiste des bases de données. Dans le second cas, qui touchait à la

passerelle OntoPop, cela a nécessité une collaboration entre le chef de produit ITM de Mondeca, la responsable de couplage de Mondeca, l'architecte innovation d'Exinis, et, dans une moindre mesure, le *core product technical manager* de Temis.

En dehors de ces points liés aux composants logiciels eux-mêmes, il s'agissait bien sûr de réaliser les définitions de l'ontologie, et de définition des RAC.

- L'ontologie a été définie par l'expert ontologue de Mondeca, qui a collaboré, en cette occasion avec le chef de projet et l'architecte innovation d'Exinis. Cela a constitué une source d'apprentissage pour Mondeca.
- Les RAC ont été définies par F. Amardeilh, qui, pour ce faire, a collaboré d'une manière très régulière avec une des rédactrices expertes de Temis, Emilie. De cette collaboration pour la définition des RAC, a émergé un enrichissement du concept de RAC, puisque celui-ci a intégré les notions nouvelles de « priorité » et de « confiance » accordée à une information. Cette évolution a été par la suite prise en compte dans le cadre du projet PressPro. Pour définir les RAC, F. Amardeilh a aussi régulièrement collaboré avec Amandine, en charge du développement informatique de la cartouche de connaissance juridique chez Temis.

De ces travaux, ont aussi émergé un besoin de consolidation lié à l'information extraite par une cartouche de connaissance et transmise à une base de connaissance : en effet, une fois la base de connaissances enrichie, il s'est avéré utile voire indispensable, d'enrichir en retour les lexiques utilisés par la cartouche de connaissance, avec cette information. Ce besoin est apparu crucial aux yeux de l'architecte innovation, et en conséquence, un nouveau composant logiciel a été développé par Mondeca, qui est venu enrichir la passerelle OntoPop.

On voit ainsi que le travail réalisé par Mondeca sur le projet Exinis, a été lourd mais riche, ce qui justifie en grande partie le retard pris par cet éditeur de logiciel dans la réalisation des tâches qui lui incombent.

### **L'application documentaire :**

La tâche de développement du système éditorial (qui correspond à l'« application documentaire » du projet PressPro) et d'intégration de la solution était dévolue à 4D Concept. Cette application documentaire consistait en une personnalisation du logiciel Epic, qui est un environnement de

rédaction en XML utilisé par les rédacteurs. Epic (logiciel packagé de l'éditeur de logiciel américain Adobe) est un atelier de rédaction semi-structuré en XML, sur lequel les auteurs vont effectuer les analyses des briques logicielles de Temis et faire les liens avec la base de connaissance de Mondeca.

Les tâches qui ont demandé le plus de travail à 4D Concept sont le développement des interfaces utilisateur, et surtout la consolidation des informations stockées au travers de l'utilisation de services web spécialisés. Ce dernier point n'avait pas été abordé lors du projet PressPro :

- Le développement des interfaces utilisateur a été réalisé par le chef de projet 4D Concept, qui a collaboré en cette occasion avec les deux rédactrices expertes et avec l'architecte innovation d'Exinis. Ce travail n'a pas posé de problème particulier.
- Par contre, la consolidation des web services, a posé un certain nombre de problèmes, et nécessité une collaboration poussée avec le chef de produit ITM de Mondeca. Ainsi, 4D Concept a dû « découper » les services web externes de consolidation, afin de pouvoir les connecter (« plugger ») à la passerelle logicielle OntoPop. Le chef de produit ITM de Mondeca, a régulièrement travaillé avec les acteurs du projet de 4D Concept sur cette question. 4D Concept a dû y consacrer beaucoup de temps, tandis que le chef de produit ITM de Mondeca a dû faire évoluer certains composants de sa base de connaissance pour permettre l'intégration de ces services web. Là encore, ce travail nouveau et difficile, a été source d'apprentissage pour les acteurs impliqués, et a contribué à faire évoluer la base ITM. C'est ce travail qui a retardé la livraison du système éditorial.

**Version V0 livrée en février 2006, la commande des licences logicielles bloquée jusqu'en juin 2006 :**

La version V0 de la solution a été livrée à Exinis en février 2006, alors qu'elle était initialement prévue pour octobre 2005. Ce retard était dû à une sous-estimation de l'ampleur du travail à réaliser, ce dont les responsables du projet chez Temis (architecte innovation et chef de projet intégration informatique) ont convenu. Mais, à l'automne 2005, le Directeur Général d'Exinis France n'a évidemment pas pu prouver à la direction internationale du groupe que le retour sur investissement du projet serait réel, le « cost saving » n'a donc pu être justifié. Ce retard a donc été à l'origine d'une certaine méfiance de la part de la direction du groupe Exinis, et a entraîné le gel de la commande de licence de production, i.e. de la commande officielle des logiciels

développés par Temis et Mondeca. Auparavant, Exinis avait commandé une licence (Temis+Mondeca) de test et développement (40 000 €) et une prestation de services (230 000 €) à répartir à part égale entre Temis et Mondeca. La recette a été réalisée fin février 2006, car toutes les briques logicielles constitutives de la version V0 étaient achevées, et leur licence ayant expiré, Exinis devait les acheter.

D'octobre 2005 à février 2006, Exinis France a toutefois pu voir progressivement les résultats de la cartouche développée par Temis, et a été impressionné par les résultats obtenus. Le CEO d'Exinis France a alors affirmé à Temis et Mondeca, que de toute façon le projet serait poursuivi en 2006, même si cela n'était plus dans le cadre du projet Kepler. Cela signifiait qu'Exinis France envisageait de le financer seul, indépendamment d'Exinis Group. Ensuite, début 2006, Temis et Mondeca ont achevé leurs prestations, et dès janvier Temis a réclamé à Exinis la commande relative à la licence d'exploitation définitive des logiciels, la « licence de production ». Mais, en mars 2006, le CEO d'Exinis France, qui soutenait le projet, a été limogé.

#### *7.5.2 Relance du projet et finalisation de la version V1 (juin 2006 - juin 2008) :*

L'achat des licences définitives des logiciels de Temis et de Mondeca était bloqué depuis huit mois, malgré l'appui du CIO France. Cela représentait théoriquement 270000 € (moitié pour Temis, moitié pour Mondeca). Pour débloquer la situation, en juin 2006, les responsables commerciaux de Temis ont décidé de faire une proposition commerciale au nom de Temis et de Mondeca, intégrant une baisse du coût des licences des logiciels. Celle-ci a été acceptée par la directrice d'Exinis Europe. Ensuite, cela a permis de débloquer cinq lignes sur le budget du projet Kepler : licence du logiciel Temis, licence du logiciel Mondeca, prestation de services de Temis, prestation de services de Mondeca, prestation de services de 4D Concept. L'objectif est alors devenu d'avoir un système en production chez Exinis France en mars 2007.

Après le projet PressPro, Temis et Mondeca se sont rendus compte qu'il fallait mettre en place certains principes, et formaliser la manière avec laquelle ils collaboraient, et ce en intégrant le client. Parce qu'il y avait eu beaucoup trop d'allers-retours, beaucoup de temps avait été perdu par ces deux éditeurs de logiciel, sans pour autant éviter une certaine incompréhension (comme entre Temis et PressPro). Cela les a incités à essayer de structurer leurs nombreuses interactions, relatives à la définition des RAC, en cohérence avec la définition d'une cartouche de connaissance et avec celle d'une ontologie : F. Amardeilh a travaillé avec les chefs de projet et les autres acteurs

concernés pour définir et cadrer ces interactions. Cela a abouti à la définition de la méthodologie OntoPop début 2006. (cf. partie 6 de ce chapitre). Les acteurs ont essayé de mettre en place cette méthodologie le plus tôt possible, à partir de juin 2006 sur le projet Exinis. La méthodologie OntoPop a donc été déployée à partir de juin 2006 chez Exinis, tandis que l'architecture logicielle qui avait été définie pour le projet PressPro, a été affinée progressivement. A partir de juin 2006, l'organisation des collaborations était donc à peu près en place.

#### **Livraison de la cartouche juridique par Temis (février 2007) :**

Temis a finalisé la version officielle de la cartouche juridique, et en a fait la livraison en février 2007, pour qu'Exinis puisse entrer dans une phase de pré-production. Les tests ont ensuite été faits par les deux rédactrices expertes d'Exinis. Par ailleurs, Exinis a demandé à Temis des développements supplémentaires concernant la cartouche, qui ont été livrés en mai 2007. A ce moment-là, il restait des problèmes d'instabilité de la base de connaissance de Mondeca à résoudre. Mais les responsables projet chez Exinis satisfaits de la contribution de Temis au projet, faisaient confiance à Mondeca pour résoudre les problèmes restants. Ensuite, Temis a travaillé sur une autre cartouche de connaissance ayant pour objectif de reprendre le contenu du Thesaurus d'Exinis, afin de réindexer, de manière homogène, cette jurisprudence. Cette tâche ne concernait pas le système éditorial.

#### **Finalisation des RAC et migration de la base Sara vers ITM (octobre 2007) :**

Mondeca a pu reconstituer la base de liens en utilisant la cartouche juridique de Temis, pour faire migrer la base de données Sara, dans la base de connaissance ITM. Les RAC ont été finalisées, et les problèmes de stabilité de la base de connaissance en grande partie résolus. Les briques logicielles de Mondeca ont ainsi pu être livrées en octobre 2007.

#### **Livraison du « système éditorial », pré-production (décembre 2007) et production (juin 2008) :**

Le système éditorial, livré par 4D Concept, est entré en pré-production en décembre 2007. Un premier groupe d'une dizaine de rédacteurs travaillant pour Exinis l'a alors testé. Les premiers retours d'information ont indiqué que le système fonctionnait, même s'il nécessitait quelques ajustements de la part de 4D Concept. Pour le reste, les améliorations à porter aux systèmes concernaient la modification de quelques patrons d'extraction par Temis, et de RAC par

Mondeca. Les craintes initiales d'Exinis sur l'instabilité de la base de connaissance se sont estompées, et la migration de l'ancienne base de données vers le nouveau système, première étape indispensable au développement de nouvelles applications, a réussi. La formation d'un certain nombre d'utilisateurs du système a ensuite été faite à la fois par 4D Concept, mais aussi par les deux rédactrices expertes. Le système est ainsi rentré en production en juin 2008, date à laquelle il restait encore à étendre l'utilisation du système à un plus grand nombre de rédacteurs. Enfin, Exinis qui souhaitait prendre le contrôle de la solution, a souhaité gérer la maintenance des RAC définies par Mondeca, ce qui a pu être réalisé en partie. Par contre, la maintenance de la cartouche de connaissance de Temis, n'a pu être transférée à Exinis.

En conclusion, les résultats obtenus par OntoPop dans le domaine de l'Édition Juridique, se sont avérés très satisfaisants, et semblent témoigner de la capacité de la passerelle OntoPop et des Règles d'Acquisition de Connaissance à s'adapter à des domaines d'application différents.

## **7.6 Synthèse - résumé : Principales différences entre les projets PressPro et Exinis.**

Les différences entre ces deux projets tiennent au fait que le projet Exinis a bénéficié des enseignements et du travail effectué par le projet PressPro, mais pas uniquement. Un certain nombre d'ingrédients dans le projet Exinis paraissent à souligner parce qu'ils ont joué un rôle dans le succès de ce projet, et qu'on ne les trouve pas dans le projet PressPro.

- Le projet s'est inscrit dans le cadre d'un vaste projet de refonte de tout le système éditorial d'Exinis, le projet Kepler, et il a eu une visibilité au plus haut niveau du groupe Exinis.
- Un intégrateur, la société de services 4D Concept, a eu en charge le développement du « système éditorial », et son intégration au système d'information d'Exinis France.
- L'architecte du projet, avait des compétences à la fois pointues et particulièrement adaptées à la compréhension des difficultés techniques du projet. Il a désigné deux rédactrices pour participer qui avaient des compétences adaptées à une participation active au projet.
- La collaboration entre Temis et Mondeca a été mieux structurée : cela résulte de ce que le couplage entre les logiciels de ces éditeurs avait déjà été réalisé pour le projet PressPro,

et de la mise en œuvre de la méthodologie OntoPop, qui a permis d'organiser les collaborations.

- Les prestataires Temis, Mondeca et 4D Concept ont été en contact direct avec de futurs utilisateurs du système d'Exinis France, pendant toute la durée du projet,
- Le travail avec les futurs utilisateurs du système a été beaucoup plus structuré : dans un premier temps avec les deux rédactrices expertes, puis au cours d'ateliers, avec les futurs utilisateurs, « guidés » par JR et ces deux rédactrices expertes.
- Exinis a eu la volonté de s'approprier la cartouche juridique, de se former et de faire ses propres développements spécifiques à terme, grâce à un *toolkit*. De même, Exinis a eu la volonté de gérer de manière autonome la maintenance des RAC sans faire appel à Mondeca.

## 8. CONCLUSIONS ET APPORTS DES PROJETS PRESSPRO ET EXINIS :

L'architecture logicielle a été définie grâce au projet PressPro, mais la définition précise de chacune des briques logicielles constitutives de la solution doit beaucoup au projet Exinis, où le niveau d'exigence était supérieur à celui du projet PressPro, du fait de la complexité du domaine étudié. Cela est aussi dû au fait que le projet Exinis a pu aller à son terme, notamment en ce qui concerne le couplage entre IDE et ITM, et la définition des RAC, à la différence du projet PressPro. D'ailleurs bien qu'il s'agisse du deuxième projet, les développements ainsi réalisés pour le projet Exinis, ont été utilisés pour la solution PressPro, car elle n'était pas encore achevée.

En ce qui concerne les RAC, le projet Exinis a apporté les notions nouvelles de « priorité » et de « confiance ». La notion de priorité peut être illustrée par l'exemple des dates de décision: il faut récupérer une date de décision dans l'arbre, et il peut y avoir dans un texte, soit une date de décision, soit une date de séance, soit une date de lecture : il faut alors fixer des priorités, suivant que la date se trouve dans la partie 1 ou 2 du document, si elle est liée à tel autre nœud de l'arbre etc., ce sont typiquement des règles liées au métier. Cette notion de priorité n'existait pas dans le projet PressPro, c'est Exinis qui l'a apporté, tout comme la notion de taux de confiance. Ainsi, Exinis a permis d'affiner les RAC. Quand le recettage a été fait chez Exinis, Emilie a constaté qu'il fallait moins d'une heure pour des modifications de RAC par Mondeca. Cela vient de ce que les RAC ont été, dès le départ, conçues pour être externalisées.

Pour Exinis, la modélisation de l'ontologie était tellement complexe, que pour pouvoir enrichir la base de connaissance, il fallait mettre en place un certain nombre de raisonnements qui étaient à la fois de l'ordre du contrôle (vérification de classe, vérification d'instances dans la base de connaissance ...) et de la réflexion par rapport au processus métier du client « il faut placer la date de décision, s'il n'y a pas eu de date de séance, etc. ». Cette distinction entre ce qui relève du métier et ce qui relève du processus général indépendamment du domaine est fondamentale pour la définition de la solution. C'est à l'occasion du projet Exinis, que Mondeca et Temis ont eu à séparer les deux, et ont essayé de définir qu'elle était la part du raisonnement métier et la part du raisonnement propre à n'importe quel système de peuplement d'ontologie. Exinis a soulevé des problèmes nouveaux, auxquels il fallait trouver une solution, sans bouleverser le cœur de l'outil, qui avait commencé à être défini dans le cadre du projet PressPro. Par exemple, il a fallu connecter un *web service* (ou service web), qui allait, dans le cadre de la consolidation, interroger un autre serveur pour obtenir des informations plus « métier », plus précises, afin de réaliser des contrôles précis. Cela a certes nécessité un travail de la part de chef de projet 4D Concept et du chef de produit de Mondeca, sur les composants de la passerelle OntoPop, mais cela illustre la capacité de la plate-forme logicielle à se connecter à des modules logiciels externes.

Les projets PressPro et Exinis ont montré que l'ontologue et le linguiste jouent des rôles distincts et concomitants, et qu'il fallait coordonner leurs tâches qui doivent être menées en parallèle. La définition d'une ingénierie s'est avérée nécessaire. Il n'est pas en effet pas envisageable que l'ontologue, après avoir défini l'ontologie du domaine, puisse tenir le rôle du linguiste et définir les patrons d'extraction. La raison en est que l'ontologue et le linguiste n'ont pas les mêmes compétences. Ontologues et linguistes ont des formations différentes, même s'il commence à apparaître des gens qui sont à la frontière des deux domaines (travaillant à la fois sur le texte et la représentation des connaissances) : un ontologue a une formation en gestion des connaissances (représentation, modélisation, graphe, réseau sémantique...), pas de linguiste. Ecrire des patrons d'extraction et définir les concepts d'une ontologie sont deux choses très différentes. Un linguiste, s'intéresse beaucoup plus à la langue elle-même, et fait différents niveaux d'analyse : morphologique, syntaxique, et sémantique. L'ontologue, lui, va surtout regarder comment fonctionne l'entreprise, et prendre en compte de ce qui existe déjà dans l'entreprise : il va reprendre tout ce qui est lié à la formalisation des bases de données et des ressources documentaires, mais il ne s'intéressera pas vraiment au texte en lui-même. L'ontologue s'intéressera aux concepts manipulés par les personnes travaillant dans l'entreprise : comment voient-ils leur métier, quels sont leurs besoins ?... Alors qu'un linguiste s'intéressera à un

ensemble de textes : de quoi parlent-ils ? Quels éléments suis-je capable d'extraire ? Comment s'articulent ces éléments entre eux ? Et finalement, que puis-je produire à partir de ces textes-là ? Ce travail de fouille de données textuelles, n'est pas le même que le travail de fouille de connaissances. L'approche est différente, même si, à la fin, on obtient des concepts identiques.

On pourrait imaginer que l'ontologue travaille d'abord, puis communique les résultats de son travail au linguiste, qui tiendrait compte alors de cette ontologie pour définir ses patrons d'extraction. Mais cela aurait plusieurs conséquences négatives. Premièrement, le linguiste serait amené à réfléchir principalement sur la base du travail de l'ontologue, c'est-à-dire sur les éléments constituant l'ontologie qu'il aurait définis (« je sais extraire telle chose, pas telle autre »). Le linguiste ne penserait alors peut-être pas à certaines choses qu'il serait capable d'extraire, et qui, à l'inverse, aurait pu servir à enrichir l'ontologie. Deuxièmement, un tel processus séquentiel prendrait sans doute plus de temps que le travail effectué en parallèle par l'ontologue et le linguiste. A la lumière de l'expérience fournie par les projets PressPro et Exinis, il apparaît donc plus facile de mettre en place un processus dialectique où ontologue et linguiste peuvent régulièrement échanger : « j'ai repéré cette chose, est-ce que tu l'as ? J'ai modélisé cet élément de l'ontologie, est-ce que tu penses qu'on peut peupler à partir de tel texte ?... ». Le linguiste analyse ce dont il dispose, et voit ce qu'il peut en extraire, tandis que l'ontologue voit ce qui existe, les besoins... Généralement, l'ontologue a une conception beaucoup plus globale, notamment parce qu'il doit prévoir les besoins ultérieurs en termes de requêtes (quelles pourraient être les requêtes formulées par les utilisateurs ?). Un troisième acteur est à associer à ce processus : le client. Il a ses besoins propres, une vision précise de son métier, mais il ne comprend pas forcément comment se conceptualisent les choses, comment il faut analyser un texte pour en ressortir des informations, il sera généralement assez « naïf » par rapport à ses questions. La méthodologie OntoPop structure des étapes dans ce dialogue à trois entre le client, le linguiste et l'ontologue.

Au terme des projets étudiés, il apparaît que l'architecture produit et l'ingénierie relative à la personnalisation et à l'intégration d'une telle solution ont été définies d'une manière relativement précise. Cette solution a été mise en production avec succès chez Exinis, où de nouveaux projets fondés sur la plate-forme associant IDE, ITM et la passerelle OntoPop ont ensuite été lancés. Temis doit encore travailler sur son environnement de développement logiciel STDK, pour qu'à terme les clients puissent maintenir et développer eux-mêmes leurs propres cartouches de connaissance. De même, Mondeca doit réfléchir aux conditions d'un transfert de compétences permettant la définition de RAC par ces mêmes clients.

Pour conclure, il est intéressant de souligner que la démarche OntoPop a été déployée ultérieurement sur les projets AFP, Thomson Scientific, et dans un autre groupe de Presse. Ainsi, les solutions développées pour les projets PressPro et Exinis ne constituent pas des applications spécifiques, mais portent en elles le noyau d'une gamme de produits que Temis et Mondeca peuvent désormais développer conjointement. Par exemple, il est possible que l'on assiste à l'avenir à une nouvelle évolution de l'outil logiciel, qui permette aussi le traitement d'images et de vidéos. La santé financière et les contrats signés par Temis et de Mondeca, au terme de notre recherche, leur permettent en effet d'envisager d'investir ces nouveaux domaines.

# **Chapitre 5 : Discussion**

## SOMMAIRE DU CHAPITRE 5 :

<b>Chapitre 5 : Discussion</b>	<b>246</b>
<i>Sommaire du Chapitre 5 :</i>	247
1. <i>Introduction :</i>	249
2. <i>Modularité :</i>	249
2.1 Deux niveaux d'innovation, architectural et local :	249
2.2 L'illusion d'une modularité d'usage :	251
2.3 Les composants de l'innovation logicielle :	252
2.3.1 Des logiciels cœur à la plate-forme de création de base de connaissance :	252
2.3.2 Les composants de la passerelle OntoPop :	254
2.3.3 L'application documentaire :	255
2.3.4 L'émergence d'une nouvelle plate-forme logicielle :	255
2.4 Conclusion sur la nature de l'offre :	256
3. <i>La structuration des projets: collaboration, coordination et compétences.</i>	258
3.1 Introduction :	258
3.2 Des collaborations interorganisationnelles multiples :	258
3.2.1 Introduction :	258
3.2.2 Les collaborations liées aux différentes tâches :	259
3.2.2.1 La passerelle logicielle OntoPop :	260
3.2.2.2 La cartouche de connaissance personnalisée.	261
3.2.2.3 L'ontologie du domaine étudié :	262
3.2.2.4 La définition personnalisée des RAC :	263
3.2.2.5 L'application documentaire :	263
3.2.2.6 Conclusion :	265
3.3 La nature des interactions entre acteurs :	266
3.3.1 Introduction :	266
3.3.2 Interactions fournisseur/client lors du projet PressPro :	267
3.3.2.1 Interaction des membres du consortium avec l'architecte innovation de PressPro:	267
3.3.2.2 Interaction entre membres du consortium et utilisateurs classiques de PressPro :	269
3.3.2.3 Interaction des membres du consortium avec les informaticiens de PressPro :	271
3.3.2.4 Interaction des membres du consortium avec le management de PressPro :	271
3.3.3 Interaction fournisseur/client lors du projet Exinis :	272
3.3.3.1 Interaction entre Temis, Mondeca et 4D Concept, et l'architecte innovation d'Exinis :	272
3.3.3.2 Interactions entre Temis, Mondeca et 4D Concept, et les utilisateurs d'Exinis :	273
3.3.3.3 Interactions entre Temis, Mondeca et 4D Concept et les informaticiens d'Exinis :	274
3.3.3.4 Interactions entre Temis, Mondeca et 4D Concept et le management d'Exinis :	274
3.3.3.5 Comparaison entre les projets PressPro et Exinis, du point de vue des interactions fournisseur/client :	275
3.3.4 Interaction entre Temis et Mondeca :	275
3.3.5 Conclusion sur la nature des interactions entre acteurs :	277
3.4 Une structuration de projet émergente :	277
3.4.1 Introduction :	277

---

3.4.2	Structuration du projet PressPro :	279
3.4.3	Structuration du projet Exinis :	282
3.4.4	Une structuration définie après une phase d'exploration :	286
3.4.5	La question de l'encastrement de la coordination :	287
4.	<i>Les rôles joués par le client dans le processus d'innovation : enrichissement de la notion de lead user.</i>	288
4.1	Introduction :	288
4.2	Le rôle du client à plusieurs niveaux :	289
4.2.1	La conception de l'architecture de l'innovation : le premier client à l'initiative du projet.	289
4.2.2	La conception des composants d'interfaçage entre les core products constituant l'innovation :	291
4.2.3	La personnalisation des logiciels core products et de la passerelle OntoPop :	292
4.2.4	La conception de l'application documentaire :	295
4.2.5	La coordination globale du projet :	297
4.3.	Synthèse des rôles joués au cours des projets PressPro et Exinis :	300
4.4	Enrichissement de la notion de <i>lead user</i> :	301
4.4.1	Des caractéristiques de lead users:	301
4.4.2	Nature des rôles joués par les acteurs clefs des projets chez PressPro et Exinis :	302
4.4.3	Enrichissement de la notion de lead user :	303
4.4.3.1	Introduction :	303
4.4.3.2	L'indispensable contribution du client au processus d'innovation :	303
4.4.3.3	Une contribution multiple du client, par des acteurs divers :	304
4.4.4	Conclusion sur la contribution du client au processus d'innovation :	307
4.5	Des tâches qui ne peuvent être assumées par le client seul :	308
5.	<i>Conclusion du chapitre :</i>	309
5.1	La modularité de l'innovation :	309
5.2	L'innovation par les utilisateurs :	310
5.3	L'apprentissage :	314

## 1. INTRODUCTION :

Le suivi de deux projets que nous avons effectué nous a permis de voir émerger une offre logicielle innovante. La première caractéristique que nous discutons est la modularité de la solution conçue, ce qui nous conduira à analyser l'innovation au niveau architectural, ainsi que le processus d'émergence de ses *design rules*. Nous allons donc dans un premier temps revenir sur la nature de l'innovation que nous étudions. Cela nous conduira à définir une typologie des logiciels constituant la solution portée par différents éditeurs de logiciel. Nous introduisons ainsi des notions nouvelles relatives à l'innovation logicielle multi-acteurs.

Ensuite, nous décrivons les collaborations interorganisationnelles mises en œuvre lors des projets étudiés. Cela nous conduit à caractériser les interactions liées à ces collaborations. Nous étudions la question de la structuration des projets étudiés, et nous décrivons comment la division du travail s'organise et comment certains rôles assumés par différents acteurs émergent progressivement.

Enfin, nous nous focalisons sur les clients, pour définir leur rôle dans le processus d'innovation. Nous discutons alors la notion de *lead user*, et proposons des catégories nouvelles pour les acteurs du processus d'innovation situés chez le client.

## 2. MODULARITE :

### 2.1 Deux niveaux d'innovation, architectural et local :

Nous considérons que l'innovation analysée est d'abord une innovation architecturale. Elle résulte de l'assemblage nouveau de logiciels développés par des éditeurs de logiciel indépendants. L'architecture modulaire, ou plutôt comme nous la nommerons « quasi-modulaire », ne résulte pas ici de la décomposition d'une architecture intégrée, et elle n'est pas précédée d'une définition précise des interfaces entre composants. Certes, les différents logiciels cœur constitutifs de la solution sont susceptibles d'être interfacés, mais cet interfaçage n'est au départ que virtuel. Ainsi, une partie du travail d'innovation consiste en l'élaboration des passerelles permettant d'interfacier ces logiciels.

L'innovation étudiée est donc imaginée initialement par le client à partir de composants logiciels indépendants que nous nommerons « quasi-modules ». Nous définissons les « quasi-modules » comme des composants logiciels qui sont à l'origine non connectables entre eux, parce que conçus indépendamment les uns des autres par différents éditeurs de logiciel, mais qui peuvent être couplés du fait de leur appartenance à un même domaine d'application (ici la gestion de données non structurées), de leur respect des normes imposées par un environnement de développement particulier (ici l'environnement Windows), et de leur compatibilité avec des formats standard (comme le format XML, langage de description de document qui facilite les échanges entre logiciels dans les secteurs de l'édition et de la documentation). Pour l'innovation que nous étudions, ces quasi-modules sont les logiciels IDE de Temis et ITM de Mondeca. La connexion entre ces quasi-modules nécessite le développement de composants logiciels qui constituent une passerelle, appelée OntoPop.

L'innovation ne se situe pas uniquement au niveau de l'architecture de cette plate-forme, elle se situe aussi au niveau des composants eux-mêmes. En effet, il faut concevoir et développer la personnalisation de certains composants logiciels, pour pouvoir les adapter au domaine d'activité du client et à sa façon de travailler, il faut :

1. Définir une ontologie et l'intégrer à la base de connaissance ITM.
2. Concevoir une « cartouche de connaissance » à interfacier avec IDE.
3. Personnaliser les RAC (« règles d'acquisition de connaissance »).
4. Définir l'interface utilisateur, i.e. une application documentaire.

Chaque client ayant des besoins particuliers, la personnalisation de ces composants que requiert ce travail doit être répétée à chaque projet. De plus, PressPro et Exinis ont fourni à Temis et Mondeca, de nouveaux champs d'application: ces sociétés n'avaient en effet pas encore déployé leur technologie dans le champ de la presse « people », et dans celui de l'édition de document juridique. Cela les a amenées à résoudre de nouveaux types de problème, ce qui a, dans le cas de Mondeca, conduit à faire évoluer, marginalement, son logiciel IDE.

Finalement, nous observons que l'ensemble constitué du logiciel IDE, de la cartouche de connaissance qui lui est couplée, de la passerelle logicielle OntoPop, du logiciel ITM, et de l'ontologie personnalisée (cf. figure 23, page 251), constitue une plate-forme, au sens donné par

Gawer et Cusumano (2002), à laquelle peuvent se connecter des logiciels, qui sont autant de modules. Ces modules sont le logiciel XML Server dans le cas du projet PressPro, et des services Web de consolidation dans celui du projet Exinis. Nous choisissons d'appeler cette plate-forme, «plate-forme de création automatisée de base de connaissance», ce qui constitue une simplification pour différentes raisons : d'abord, parce que le caractère « automatique » des opérations n'est en fait parfois que « semi-automatique », l'intervention humaine étant nécessaire ; ensuite, parce que le terme « création » devrait être précisé par ceux « d'enrichissement et de maintenance » ; enfin, parce qu'il faudrait ajouter la notion d'annotation de documents permise par une telle solution.

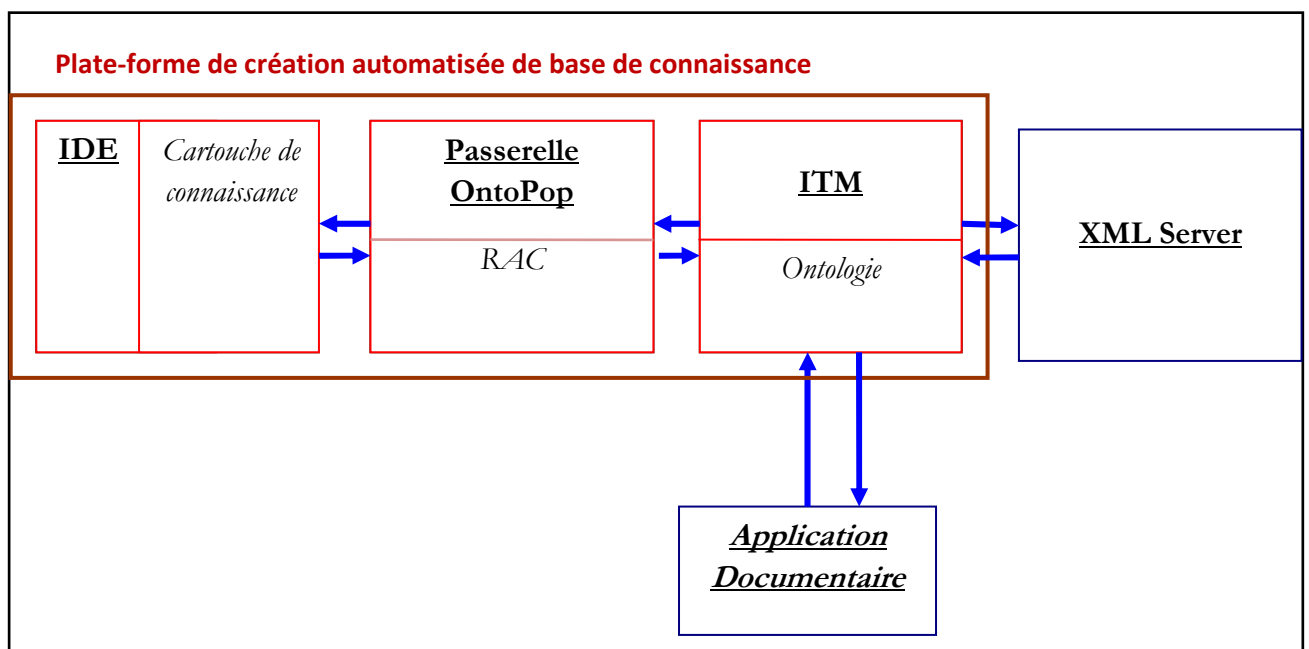


Figure 23 : Les composants [Modules (soulignés) et Composants non modulaires] formant la solution PressPro, incluant la plate-forme de « création automatisée de base de connaissance » [les composants personnalisés sont indiqués en italiques].

## 2.2 L'illusion d'une modularité d'usage :

Notre recherche met en évidence une première originalité notable par rapport aux enseignements de la littérature sur la modularité. Les composants logiciels, que nous définissons comme « quasi-modules », précèdent la définition de l'innovation qui les intègre. Cela constitue une des caractéristiques de l'industrie du logiciel, que de proposer des produits qui peuvent être combinés pour aboutir à une offre nouvelle, mais qui sont des logiciels autonomes, tant du point de vue

technique, car ils peuvent fonctionner seuls, que du point de vue marché, car ils peuvent être commercialisés séparément.

Les logiciels développés pour un même environnement informatique, nous l'avons dit, peuvent théoriquement être reliés les uns aux autres. Concrètement, nous observons que cela peut créer l'illusion d'une modularité d'usage, comme cela a été le cas pour le premier client. En effet, ce client a sous-estimé plusieurs difficultés : celle d'intégrer les logiciels choisis à une même solution, mais aussi le caractère radicalement innovant de certains de ces composants. Il ne s'agissait donc pas de composants standardisés prêts à être combinés pour former une application rapidement opérationnelle et utilisable, comme on l'observe en modularité d'usage dans des domaines comme la microinformatique. Mais c'est bien grâce à cette illusion de modularité d'usage (Ulrich 1995 ; Baldwin et Clark, 2000), que le projet PressPro a pu être lancé, et donc que l'innovation a pu être conçue.

### **2.3 Les composants de l'innovation logicielle :**

Nous précisons maintenant la « nature » des composants de l'innovation. Celle-ci, au terme de ces projets, se compose de briques logicielles, qui sont de différents types : il y a tout d'abord les « quasi-modules » initiaux, c'est-à-dire des composants logiciels susceptibles d'être connectés les uns aux autres ; ensuite, il y a des composants logiciels développés spécialement pour relier ces quasi-modules, formant une « passerelle » logicielle. Ensuite, des composants logiciels et des paramétrages sont réalisés pour personnaliser les quasi-modules initiaux. Enfin, nous trouvons l'application documentaire. Nous allons maintenant revenir plus en détail sur la nature de ces différents composants.

#### *2.3.1 Des logiciels cœur à la plate-forme de création de base de connaissance :*

L'innovation est fondée sur des composants logiciels, qui font partie des logiciels « cœur » des éditeurs de logiciel Temis et Mondeca. Ces logiciels supposent pour fonctionner une personnalisation réalisée :

- soit par un composant logiciel de personnalisation : pour Temis, il s'agit du logiciel IDE, le moteur d'extraction d'information, dont la personnalisation est réalisée grâce à une cartouche d'extraction terminologique appelée de manière générique « Skill Cartridge »,

---

ou cartouche de connaissance, qui doit être personnalisée en définissant des concepts (« personnalités ») à extraire, ainsi que les relations entre ces concepts (« mariage »), et ce au travers de la définition de « règles d'extraction », règles qui doivent être définies par un linguiste. Ce linguiste doit bien évidemment pour cela travailler avec le client, dont il ignore a priori les usages, spécialement lors du premier déploiement d'un tel logiciel dans un secteur donné, le service documentation d'un groupe de presse « people » dans le cas PressPro.

- soit une couche de personnalisation directement intégrée dans la brique logicielle : pour Mondeca, il s'agit du logiciel ITM, système de représentation de connaissance. Il s'agit ici de concevoir une ontologie, en collaboration avec le client, c'est-à-dire une représentation organisée de la connaissance propre à un domaine, et de l'intégrer à la base de connaissance ITM.

Dans le cadre de la plate-forme de création de base de connaissance, les quasi-modules initiaux, IDE et ITM, deviennent ainsi des modules au sens propre du terme, i.e. des composants de la solution qui ne sont pas affectés par les évolutions éventuelles des autres composants. Par contre, la cartouche de connaissance et l'ontologie constituent des composants personnalisés, qui ne sont pas des modules. En effet, la cartouche de connaissance, l'ontologie mais aussi les RAC sont interdépendantes : la modification de l'un de ces éléments implique a priori de modifier les deux autres.

### 2.3.2 Les composants de la passerelle OntoPop :

Les logiciels qui constituent la passerelle logicielle OntoPop sont les suivants (figure 24) :

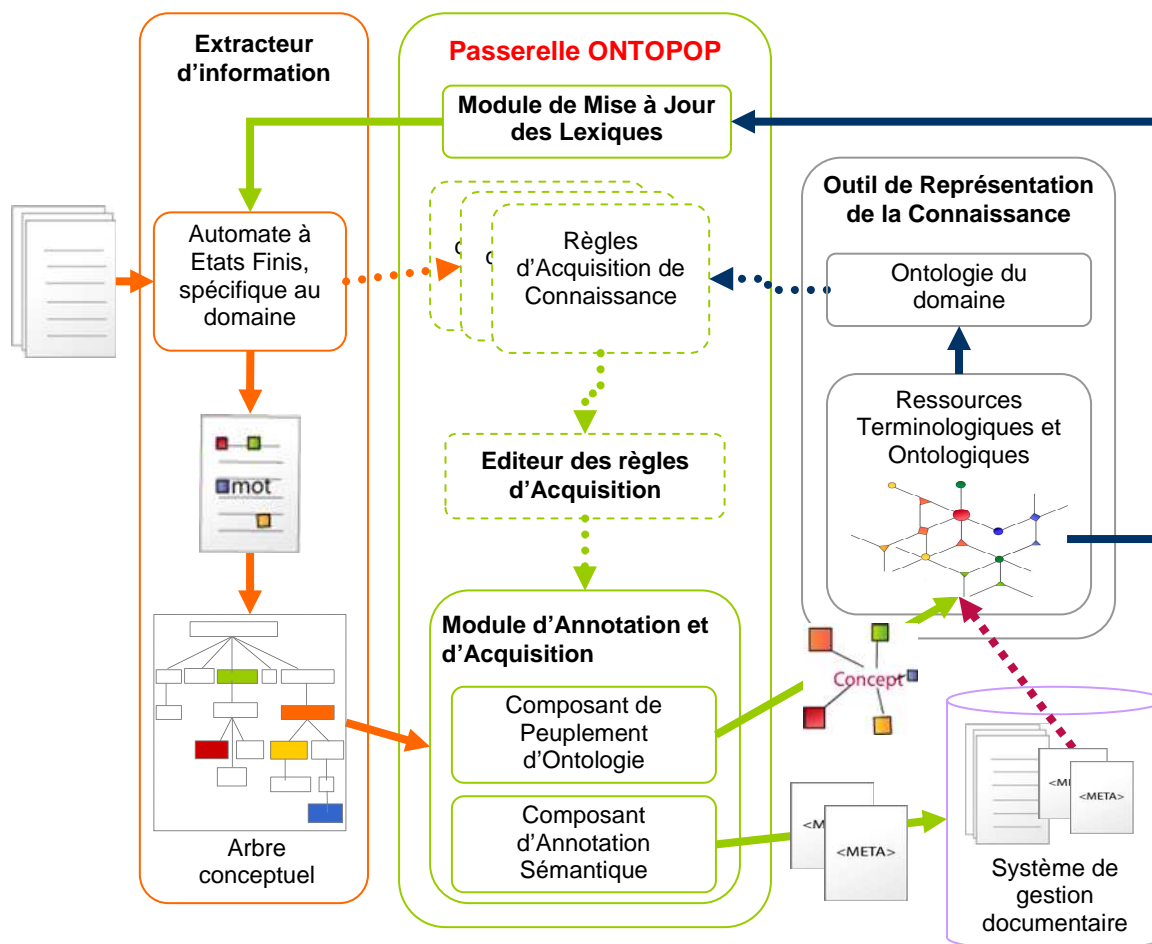


Figure 24 La passerelle logicielle OntoPop (D'après Amardeilh, 2007).

- **L'éditeur de « règles d'acquisition de connaissance » (« RAC »)**, qui permet de saisir l'ensemble des règles d'acquisition de connaissance. Cet éditeur de RAC est un module, qui se connecte au logiciel d'annotation et d'acquisition.
- **Le composant d'annotation et d'acquisition** : il est lui-même composé du module d'extraction d'information (interfacé avec un moteur d'extraction d'information, en l'occurrence IDE ou GATE), du module de peuplement d'ontologie (acquisition de connaissance et enrichissement terminologique), du module d'annotation sémantique (permettant la normalisation terminologique), du module de stockage, et enfin de l'éventuelle interface de validation. Il s'agit d'un module, mais qui intègre une couche logicielle paramétrée via la définition des RAC.

- Enfin, **un module de mise à jour ou de maintenance des lexiques** : ce module permet de capitaliser sur les corrections et validations rentrées par les utilisateurs. Pour cela, il envoie au moteur d'extraction d'information, les libellés des nouvelles instances ou de nouveaux descripteurs du référentiel.

### 2.3.3 L'application documentaire :

L'application documentaire s'interface directement avec ITM, et indirectement avec IDE au travers du logiciel ITM. Cette application doit aussi prendre en compte les éventuels modules qui viennent compléter la plate-forme de création de base de connaissance. Elle doit être définie à chaque nouveau projet, en fonction des usages et besoins de l'entreprise cliente. Par ailleurs, elle constitue un module, dans le sens où, après qu'elle a été développée, des modifications ultérieures portant sur les autres logiciels de la solution ne l'affecteront pas.

### 2.3.4 L'émergence d'une nouvelle plate-forme logicielle :

Le couplage entre le logiciel ITM et le moteur d'extraction IDE, constitue la fondation de la plate-forme applicative, susceptible d'accueillir différents types de logiciel comme l'indique la figure ci-dessous (figure 25) :

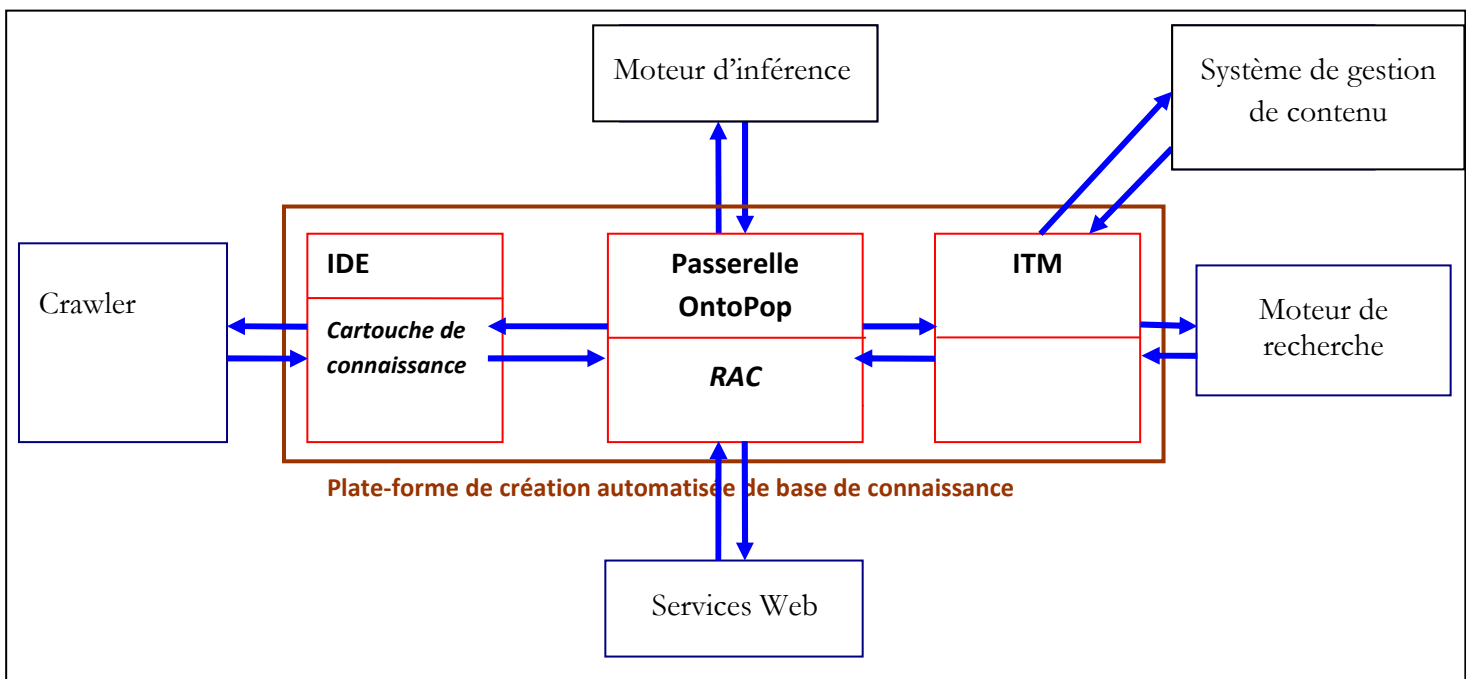


Figure 25 : Exemples connexions de différents types de logiciel à la plate-forme de création automatisée de base de connaissance.

---

Nous avons vu qu'il était en effet possible de connecter à la plate-forme de « création automatisée de base de connaissance », des systèmes de gestion de contenu (Content Management System ou CMS) et des services Web. Mais cela concerne aussi d'autres types de logiciel : crawler (robot d'indexation qui collecte des sources pour alimenter le moteur d'extraction), moteur de recherche (qui permet d'effectuer des recherches dans la base de connaissance), moteur d'inférence (qui permet d'appliquer des raisonnements lors du processus d'acquisition de connaissance). Ces différentes possibilités ont été mises en œuvre dans des projets ultérieurs par Mondeca et Temis

## 2.4 Conclusion sur la nature de l'offre :

Le processus d'innovation étudié a eu pour point de départ une architecture qui n'était pas « pré-modulaire » au sens de Baldwin et Clark (2000), qui attribue ce qualificatif aux architectures produit intégrées, susceptibles d'être dé-intégrées, i.e. découpée en modules. Nous proposons le terme d'architecture « **quasi-modulaire** », pour désigner l'état initial de cette architecture. Nous définissons une architecture logicielle quasi-modulaire comme une architecture logicielle réunissant des logiciels « cœur » indépendants les uns des autres, dont l'interconnexion nécessite le développement de passerelles logicielles, certaines génériques, d'autres personnalisables. Ces différentes personnalisations peuvent être interdépendantes, mais leur degré d'interdépendance, n'est pas forcément connu au départ. Le processus de mise en œuvre de cette nouvelle architecture produit conduit *in fine* à déterminer parmi les différents composants logiciels qu'elle réunit, ceux qui constitueront des modules et ceux qui resteront dépendants de composants portés par d'autres éditeurs de logiciel. Pour ces derniers nous proposons la terminologie de « **pseudo-module** », afin de distinguer ces composants des modules et des quasi-modules. Le processus de mise en œuvre permettra aussi de distinguer les composants génériques, donc réutilisables tel quel, des composants ad hoc, qui devront être redéveloppés à chaque nouveau projet mettant en œuvre l'innovation étudiée. Cela nous permet de décrire quatre types de composants logiciels, constitutifs du type d'innovation logicielle multi-acteurs, comme l'indique le tableau 4, page suivante.

	Module	Pseudo-module
<b>Composant générique</b>	- Logiciels IDE et ITM, - Passerelle OntoPop, - Logiciels composants la passerelle logicielle OntoPop, - XML server (projet PressPro).	
<b>Composant ad hoc (ou paramétrage de composant générique).</b>	- Application documentaire.	- Cartouche de connaissance, - Ontologie, - RAC (règles d'acquisition des connaissances).

Tableau 4 : Typologie des composants logiciels d'une innovation logicielle multi-acteurs.

In fine, cette plate-forme prise comme un tout constitue bien un module au sens propre du terme, mais comme elle n'est pas constituée que de modules, nous la qualifierons de **plate-forme « quasi-modulaire »** eu égard au fait qu'elle a été conçue à partir de quasi-modules, et qu'elle n'est pas encore exclusivement composée de modules. Ainsi, la solution est très proche d'être modulaire, mais elle ne l'est pas tout à fait. Les trois composants non modulaires de cette plate-forme (cartouche de connaissance, ontologie, et RAC) sont donc des « **pseudo-modules** », au sens que nous venons de définir.

Nous représentons sur la figure 26, la solution développée pour PressPro, en distinguant les modules des pseudo-modules, et les composants génériques des composants ad-hoc :

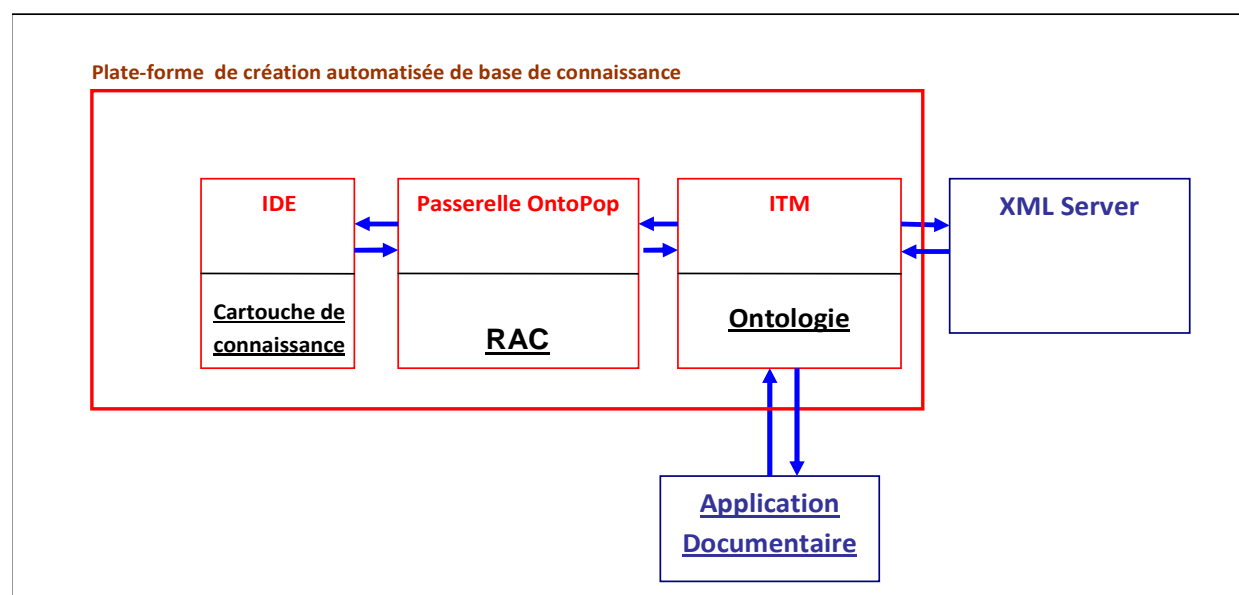


Figure 26 : Architecture de la solution développée pour Press Pro :

- les modules de la plate-forme sont en rouge, et ses pseudo-modules en noir.
- Les modules qui se connectent à la plate-forme sont en bleu.
- les composants ad-hoc (ou personnalisés) sont soulignés.

Comme l'indique la figure 26 (page précédente), la plate-forme de « création automatisée de base de connaissance » peut s'interfacer avec l'application documentaire, et avec tout autre composant modulaire, comme par exemple le logiciel de gestion de contenu, XML Server.

La conception de l'offre qui a émergé des projets PressPro et Mondeca, a nécessité la mise en place d'une série de collaborations entre firmes indépendantes. Ces collaborations ont abouti à la conception, et à la détermination des conditions de mise en œuvre, de l'innovation. Nous allons maintenant analyser ces différentes collaborations.

### **3. LA STRUCTURATION DES PROJETS: COLLABORATION, COORDINATION ET COMPETENCES.**

#### **3.1 Introduction :**

Le travail effectué au cours des projets PressPro et Exinis, passe par la mise en œuvre de collaborations entre différentes organisations, qui travaillent ensemble pour la première fois. Ces collaborations génèrent des nombreuses interactions entre acteurs, que nous caractérisons à l'aide du « modèle d'interaction », défini par des chercheurs en marketing interorganisationnel (Hakansson, 1982 ; Valla, 1987). Nous abordons ensuite la question de la coordination des différentes collaborations menées dans le cadre de ces projets. Ces différentes analyses nous permettent de tirer des conclusions relatives aux modalités d'organisation des projets de mise en œuvre de l'innovation étudiée, et de mettre en évidence les rôles qui doivent être assumés au sein de ces projets. Enfin, nous soulignons les compétences requises pour endosser de tels rôles.

#### **3.2 Des collaborations interorganisationnelles multiples :**

##### *3.2.1 Introduction :*

Cinq tâches sont nécessaires au développement de la solution globale. A ces tâches s'ajoutent la définition de l'architecture de l'innovation et l'intégration informatique de la solution :

1. Conception et développement IDE et ITM, au travers de la conception et du développement de la passerelle OntoPop,
2. Définition de la cartouche de connaissance personnalisée,

3. Définition de l'ontologie du domaine étudié,
4. Définition des « règles d'acquisition des connaissances », permettant le couplage du logiciel IDE associé à la cartouche de connaissance, et du logiciel ITM, intégrant l'ontologie,
5. Conception et développement de l'application documentaire.

Chacune de ces tâches nécessite des collaborations interorganisationnelles. Nous allons maintenant décrire ces différentes collaborations à l'œuvre dans le cadre des projets PressPro et Exinis.

### 3.2.2 Les collaborations liées aux différentes tâches :

Nous reprenons dans le tableau 5 ci-après, les différents rôles qui sont apparus dans les projets étudiés, afin de stabiliser un vocabulaire, alors que les appellations des acteurs ont été contingentes aux entreprises dans lesquelles les applications ont été développées :

Entreprise	Fonction dans l'entreprise	Rôle dans le(s) projet(s) étudié(s)
<b>PressPro</b>	Chef de documentation	Architecte innovation
	Responsable informatique de la documentation	Chef de projet intégration informatique
	Documentalistes	Utilisateurs classiques
<b>Exinis</b>	Informaticien de la DSI, formé à l'analyse sémantique	Architecte innovation
	Informaticien de la DSI spécialiste des bases de données	Chef de projet intégration informatique
	Deux rédactrices juridiques, formées au management des connaissances	Utilisateurs experts
	Groupe de rédacteurs (éditeurs et auteurs) juridiques	Utilisateurs classiques
<b>Temis</b>	Linguiste expert	Chef de projet cartouche de connaissance
	Chef de projet/linguiste informaticien	Chef de projet Temis
<b>Mondeca</b>	Ontologue expert	Chef de projet ontologie
	Chef de projet OntoPop/responsable intégration IDE-ITM.	Chef de projet couplage
	Chef de projet/Informaticien formé à la gestion des connaissances	Chef de projet Mondeca

<b>Xyleme [projet PressPro, seulement].</b>	Directeur technique/chef de projet.	Chef de projet Xyleme
<b>4D Concept [projet Exinis, seulement].</b>	Consultant	Chef de projet 4D Concept, responsable de l'application documentaire
<b>Unilog [projet Exinis, seulement].</b>	Consultant	Chef de projet Unilog

**Tableau 5 : Principaux acteurs des projets PressPro et Exinis (fonctions dans l'entreprise, rôle dans le ou les projets).**

### 3.2.2.1 La passerelle logicielle OntoPop :

#### Les composants logiciels de la passerelle OntoPop :

- **Dans le cadre du projet PressPro** : La passerelle OntoPop a été développée principalement dans le cadre du projet PressPro, et cette tâche n'a pas eu à être renouvelée par la suite. Cette brique et ses différents composants ont été conçus par la chef de projet couplage chez Mondeca, qui a travaillé sur le sujet avec le linguiste expert de Temis, et l'ontologue expert de Mondeca. Le développement informatique en a été réalisé par deux informaticiens de Mondeca à partir des spécifications définies par la responsable de couplage. Le client PressPro n'a pas joué de rôle direct dans cette tâche.

- **Dans le cadre du projet Exinis** : L'architecte innovation d'Exinis, est intervenu pour inciter au développement d'une plate-forme plus performante, après avoir fait développer en interne une base de connaissance virtuelle. Des tests réalisés par les informaticiens d'Exinis ont mis en évidence des problèmes de lenteur au niveau du couplage IDE/ITM, et des problèmes de stabilité de la base de connaissance ITM. L'architecte innovation a fait remonter ces deux problèmes, qui concernaient le chef de projet Mondeca et la chef de projet couplage de Mondeca. Cela les a conduits à faire évoluer certains composants du logiciel ITM et de la passerelle OntoPop. D'autre part, une rédactrice d'Exinis, dotée de compétences particulières, a travaillé avec la chef de projet couplage de Mondeca. Elle était l'une des deux documentalistes, que nous qualifions d'« utilisateurs experts ». Le choix de la dénomination « utilisateurs experts », vient de ce que ces rédactrices ont reçu, en plus de leur formation de juriste, une formation dans le domaine de l'ingénierie documentaire et de la représentation des connaissances, sans en être toutefois des spécialistes. A ce titre, elles ont apporté leur expérience de rédactrice juridique, tout en comprenant et en adhérant à la vision de l'architecte innovation. L'utilisateur expert collaborant avec la chef de projet couplage de Mondeca, a donc demandé à ce que des fonctions de consolidation soient ajoutées à la solution. Cette demande a été appuyée par l'architecte innovation, et Mondeca a développé un « module de maintenance des lexiques », qui permet

d'effectuer des opérations de consolidation des ressources linguistiques. Ce composant a ensuite été incorporé à la passerelle OntoPop.

### La conception des RAC :

- **Dans le cadre du projet PressPro :** La notion de RAC a été définie par la chef de projet couplage de Mondeca, assistée du linguiste de Temis et de l'ontologue de Mondeca.
- **Dans le cadre du projet Exinis :** La notion de RAC a été enrichie, en intégrant les notions complémentaires de « priorité » et de « confiance ». Cet enrichissement résulte là encore des observations faites par un des deux utilisateurs experts collaborant de manière régulière avec la chef de projet couplage, et confirmées par l'architecte innovation d'Exinis. Ces options ont été intégrées de manière générique au concept de RAC.

#### 3.2.2.2 La cartouche de connaissance personnalisée.

- **Dans le cadre du projet PressPro :** Elle a donné lieu à l'organisation d'ateliers utilisateurs qui n'ont fonctionné que quelques mois, et ont rassemblé une demi-douzaine d'utilisateurs classiques, l'architecte innovation de PressPro, le chef de projet et le linguiste expert de Temis. Après l'arrêt de ces ateliers, les collaborations relatives à la cartouche de connaissance n'ont plus concerné que le chef de projet et le linguiste expert de Temis, d'une part, et l'architecte innovation de PressPro, d'autre part, jusqu'à ce que les autres briques logicielles soient au point. Enfin, dans un troisième temps et pendant une période assez courte, cette collaboration a concerné le chef de projet de Temis, la nouvelle responsable de documentation et un certain nombre d'utilisateurs classiques de PressPro. Ces derniers testèrent la cartouche de connaissance que Temis avait développée, et aboutirent rapidement à son rejet définitif. Un des enseignements du projet PressPro est que le développement interactif de la cartouche de connaissance entre Temis et le client, est une tâche qui peut, en partie (développement et test), être conduite séparément des autres.
- **Dans le cadre du projet Exinis :** des échanges entre le comité de projet, les deux utilisateurs experts et un groupe d'« utilisateurs classiques » ont eu lieu, en début de projet, afin d'analyser leur façon de travailler. Nous retiendrons cette dénomination d'« utilisateurs classiques », pour les distinguer de leurs collègues, utilisateurs experts. Ces deux utilisateurs experts, justement, ont ensuite travaillé et défini ce qu'il fallait extraire avec le linguiste expert de Temis. Par contre, la collaboration entre Temis et les utilisateurs classiques, « non experts », fut inexistante. Concernant la cartouche de connaissance, l'apport des utilisateurs experts fut donc déterminant,

celui des groupes d'utilisateurs classiques faible, en dehors de leur participation initiale à quelques ateliers.

### 3.2.2.3 L'ontologie du domaine étudié :

- **Dans le cadre du projet PressPro :** La création d'une base de connaissance nécessite le travail d'un expert en représentation des connaissances, l'ontologue expert, qui doit s'appuyer sur des spécialistes de ce secteur, en l'occurrence la presse « people », pour concevoir l'organisation des connaissances relatives à ce domaine. Il s'agit donc dans un premier temps pour cet ontologue de recueillir toutes les informations utiles auprès de ceux-ci. C'est ainsi que l'ontologue expert de Mondeca a travaillé lors d'ateliers, avec des groupes d'utilisateurs classiques, et ce jusqu'à leur arrêt brutal. Ensuite, il a travaillé avec les deux utilisateurs classiques dédiés à cette tâche et travaillant dans les locaux de Temis. La priorité donnée à cette tâche venait de la volonté de PressPro de disposer aussi rapidement que possible d'une solution susceptible de faire au moins ce que proposait l'ancien système. Au-delà de la compréhension de la connaissance liée au secteur, il s'agissait aussi pour Mondeca, de définir une ontologie adaptée au mode de travail des futurs utilisateurs. Ainsi, pour définir une ontologie, le chef de projet ontologie a besoin de savoir comment les utilisateurs travaillent, quel type de connaissance ils recherchent, et comment ils interrogent la base de connaissance. C'est pourquoi la compréhension précise des usages des utilisateurs est indispensable à la définition de l'ontologie.

- **Dans le cadre du projet Exinis :** Pour la conception de l'ontologie, le chef de projet ontologie de Mondeca a travaillé principalement avec le chef de projet intégration informatique, spécialiste des bases de données, et les deux utilisateurs experts d'Exinis. Il y a donc eu des réunions très régulières. Ensuite, et à la différence de la tâche de réalisation de la cartouche de connaissance, des utilisateurs classiques étaient ponctuellement sollicités pour formuler un avis au cours d'ateliers utilisateurs, en fait, implicitement, pour valider les options retenues par le chef de projet intégration informatique et les rédacteurs experts d'Exinis. Il s'agissait aussi, grâce à ses utilisateurs, de convaincre progressivement les utilisateurs classiques de l'intérêt de l'innovation en cours de développement, de les « évangéliser » selon le terme utilisé dans l'industrie informatique. Un des enseignements majeurs des projets PressPro et surtout Exinis sur cette question est que la prise en compte des usages des utilisateurs classiques est déterminante dans la définition de l'ontologie, et plus importante que pour la définition d'une cartouche de connaissance.

### 3.2.2.4 La définition personnalisée des RAC :

Cette tâche est liée à deux tâches concourantes (définitions de la cartouche de connaissance et de l'ontologie) menées par des organisations différentes. De plus c'est une tâche récurrente, qui doit être effectuée à chaque projet, et nécessite un chef de projet couplage.

- **Dans le cadre du projet PressPro** : Le processus de définition des RAC n'a pas été achevé, mais il a été engagé au travers d'une collaboration entre la chef de projet couplage, le chef de projet cartouche de connaissance et le chef de projet ontologie. Du côté du client, c'est l'architecte innovation qui a apporté sa contribution, limitée, à cette tâche. Mais finalement, les utilisateurs classiques n'ont pas réellement testé les RAC ainsi définies, le projet ayant été suspendu.

- **Dans le cadre du projet Exinis** : La personnalisation des RAC a été réalisée par la chef de projet couplage travaillant avec un utilisateur expert. Cette collaboration, supervisée par l'architecte innovation d'Exinis, a permis la conception de la partie optionnelle des RAC, prenant en compte les notions de confiance et de priorité.

### 3.2.2.5 L'application documentaire :

- **Dans le cadre du projet PressPro** : Elle devait être développée par deux informaticiens de PressPro, sous la direction du chef de projet intégration informatique, après une rapide formation par respectivement Temis, Mondeca et Xyleme. Ces informaticiens ont échoué, ne parvenant pas à débiter le travail relatif à cette tâche. PressPro ne voulait pas d'intervention de sociétés tierces, comme les sociétés de services informatiques, et ce pour plusieurs raisons :

1. Tout d'abord, les compétences en *text mining* étant très rares, ces sociétés risquaient de ne pas maîtriser ces technologies nouvelles, et en conséquence de faire échouer le projet. En effet, aucune SSII identifiée par PressPro ne possédait de connaissance relative à ces technologies nouvelles, ni aux logiciels concernés.
2. Ce manque de maîtrise les auraient placées dans une position inconfortable pour convaincre les utilisateurs finaux du bien fondé de l'innovation lors des ateliers utilisateurs ; ce rôle pédagogique (d'« évangelisation » selon les termes du chef de projet de Temis), était important aux yeux de PressPro et des membres du consortium,

3. Le fait d'ajouter un acteur risquait de compliquer la gestion du projet dans son ensemble, et de rendre son contrôle moins aisé.

Mais PressPro manquait d'un minimum de maîtrise des technologies employées, et, à un degré moindre, d'expérience en gestion de projet logiciel, tandis que les membres du consortium ne voulaient pas assumer cette tâche, car ils considéraient que cela ne faisait pas partie de leur « métier ». Quand PressPro par l'intermédiaire de son chef de projet intégration informatique et de deux informaticiens a tenté de développer seul l'application documentaire, une formation par les membres du consortium à leurs logiciels respectifs a eu lieu, mais de manière ponctuelle, i.e. une journée par logiciel « cœur ». Cela ne leur a pas suffi pour aborder efficacement cette tâche, à la surprise des chefs de projet du consortium.

Ensuite, l'architecte innovation de PressPro a demandé aux chefs de projet du consortium de prendre en charge cette tâche. Elle fut finalement attribuée à Xyleme, assisté de Mondeca dans un premier temps. Ce choix, venait de ce que Xyleme et Mondeca étaient plus proches que Temis des utilisateurs à ce moment-là. Xyleme s'est acquitté de cette tâche malgré les difficultés dues à son manque de moyens humains, à son inexpérience quant au développement de telles applications, tâche qui relève de l'intégration informatique, ainsi qu'à sa méconnaissance des logiciels de Temis et de Mondeca. Cette tâche requiert le concours des utilisateurs finaux, car il s'agit d'élaborer l'interface utilisateur de l'application documentaire qu'ils seront amenés à utiliser. Cette tâche requiert aussi des échanges avec Mondeca, car l'application documentaire s'interface directement avec la base de connaissance. Ainsi, la rédaction des spécifications de l'application documentaire a nécessité une collaboration régulière entre le chef de projet de Xyleme et les deux utilisateurs classiques dédiés de PressPro, d'une part, et des échanges ponctuels entre le chef de projet de Xyleme et le chef de projet ontologie de Mondeca, d'autre part.

- **Dans le cadre du projet Exinis** : Les collaborations interorganisationnelles ayant conduit à spécifier et à développer l'application documentaire pour Exinis, ont été les suivantes :

1. Les deux utilisateurs experts, le chef de projet intégration informatique et l'architecte innovation d'Exinis, qui se réunissaient régulièrement avec le chef de projet de 4D Concept,

2. Les deux utilisateurs experts et l'architecte innovation d'Exinis, et des groupes d'utilisateurs classiques d'Exinis à qui l'on demandait, ponctuellement, de choisir entre plusieurs propositions.
3. Le chef de projet de 4D Concept et le chef de projet ontologie de Mondeca, pour la coordination application documentaire/ontologie, assistés du chef de projet intégration informatique d'Exinis.

En conclusion sur la réalisation de l'application documentaire, nous observons qu'elle requiert une prise en compte des usages des utilisateurs aussi poussée que pour la définition de l'ontologie. Nous constatons, par ailleurs, qu'elle nécessite une coordination entre le chef de projet application documentaire et le chef de projet ontologie, mais beaucoup moins forte, que la coordination nécessaire à la définition des RAC.

#### 3.2.2.6 Conclusion :

L'analyse des projets permet de mettre en évidence les différents types de collaboration interorganisationnelle à instaurer pour développer la plate-forme de création automatisée de base de connaissance. Elle montre que chacune des tâches à réaliser nécessite, à des degrés divers, de telles collaborations. Ainsi c'est l'ensemble du processus d'innovation, dont les collaborations intra-organisationnelles (en particulier, le développement de la passerelle OntoPop, des RAC et de l'ontologie chez Mondeca, et celui de la cartouche de connaissance chez Temis), qui est subordonné à ces collaborations interorganisationnelles. Les organisations impliquées n'ayant jamais collaboré auparavant, les projets PressPro et Exinis ont permis de faire émerger progressivement la nature et les modalités de ces collaborations, dans un contexte de faible couplage interorganisationnel. L'étude de ces collaborations nous permet d'identifier différents rôles joués par les éditeurs de logiciel, le client, et les sociétés de services. Ces collaborations génèrent des interactions entre acteurs, interactions que nous allons maintenant chercher à caractériser.

### 3.3 La nature des interactions entre acteurs :

#### 3.3.1 Introduction :

De nombreuses interactions entre acteurs d'organisations différentes ont été nécessaires pour mener à bien les projets étudiés. La caractérisation de ces interactions grâce au « modèle d'interaction », issu des travaux du groupe IMP en marketing interorganisationnel (Hakansson, 1982), vise à une meilleure compréhension du déroulement des projets PressPro et Exinis, et du rôle de ses acteurs. Ce modèle d'interaction, qui concerne les relations interorganisationnelles, est destiné à l'analyse des relations fournisseur/client, mais peut aussi être utilisé pour l'analyse des relations entre fournisseurs participant à la conception et au développement d'une offre (Hakansson, 1987).

Les travaux du groupe IMP sur l'interaction fournisseur/client (Hakansson, 1982 ; Valla, 1987) soulignent le rôle crucial joué par les interactions fournisseur/client et entre fournisseurs dans les processus d'innovation (Hakansson, 1987, 1989). Dans ces conditions, il est important que les fournisseurs concernés, comprennent les caractéristiques de ces interactions afin d'optimiser la gestion de ces processus. L'étude de ces interactions nous permet de souligner le rôle qu'y jouent les compétences et la motivation des acteurs, et donc de la nécessité pour les fournisseurs de rechercher ou de développer les compétences des clients, ainsi que des sociétés tierces, comme les SSII.

Les interactions entre acteurs ont été nombreuses pour plusieurs raisons : parce que de nombreuses collaborations ont été nécessaires à la conception de la solution recherchée, mais aussi parce que ces collaborations étaient entièrement nouvelles, pour la plupart, et que leur objet aussi était entièrement nouveau : Temis et Mondeca n'avaient jamais travaillé ensemble, ni conçu de solution couplant moteur d'extraction et système de représentation des connaissances. De même, PressPro n'avait jamais travaillé avec aucun membre du consortium, tout comme Exinis n'avait jamais travaillé avec Temis, ni Mondeca. La seule exception, concerne Exinis et 4D Concept. Chaque acteur a donc été amené à découvrir les autres acteurs, en même temps que le problème technique à résoudre. Une autre raison expliquant le grand nombre d'interactions, vient de la volonté farouche de Temis et Mondeca de mener à bien ces projets, dans la perspective d'acquérir des références significatives et d'ouvrir un nouveau marché.

---

Nous allons donc analyser les interactions entre acteurs des projets PressPro et Exinis, en retenant certains des items proposés par le « modèle d'interaction », en fonction de leur capacité à caractériser les interactions observées. Nous avons ainsi choisi les items liés aux enjeux, motivations et compétences des différents acteurs, à l'atmosphère de leurs différentes relations, ainsi qu'au pouvoir relatif de ces acteurs au sein de leur relation. Pour l'analyse des interactions entre Temis et Mondeca, nous retenons aussi l'item « environnement général », qui détermine, en partie, leur nature.

### 3.3.2 Interactions fournisseur/client lors du projet PressPro :

#### 3.3.2.1 Interaction des membres du consortium avec l'architecte innovation de PressPro:

Ces interactions ont concerné d'un côté les chefs de projet de Temis, Mondeca et Xyleme, le chef de projet cartouche de connaissance de Temis et le chef de projet ontologie de Mondeca, et, de l'autre, l'architecte innovation et le chef de projet intégration informatique de PressPro. Elles furent nombreuses et régulières pendant pratiquement toute la durée du projet, et ont évolué en fonction des trois phases de ce projet :

1. Première phase : elle se caractérise par un climat de grande confiance, voire même de complicité. Celui-ci est né du sentiment de participer à un projet radicalement novateur et du sentiment d'avoir à affronter ensemble les réticences des utilisateurs classiques.
2. Deuxième phase : à partir de décembre 2004, deux problèmes se sont posés. Tout d'abord, le refus des utilisateurs classiques de poursuivre les ateliers utilisateurs, et d'autre part les difficultés rencontrées par les informaticiens de PressPro pour développer l'application documentaire. A partir de ce moment-là, la confiance des membres du consortium dans les compétences des acteurs du projet chez PressPro a décliné, et les briques logicielles de Xyleme et de Mondeca ont été jugées prioritaires par le client. Les interactions de l'architecte innovation avec Mondeca et Xyleme furent donc plus nombreuses qu'avec Temis. Cette société a alors travaillé d'une manière plus autonome au développement de la cartouche de connaissance « people », maintenant des contacts ponctuels avec l'architecte innovation de PressPro. Le climat de la relation entre Temis et l'architecte innovation restait bon. Par contre, le climat s'est tendu entre l'architecte innovation et les chefs de projet de Mondeca et de Xyleme.

3. Troisième phase : à partir de fin 2005, phase de pré-production, un comité de qualité étant en place chez PressPro, les relations se sont progressivement dégradées entre Temis et l'architecte innovation de PressPro. Les difficultés techniques concernant les briques de Mondeca et Xyleme, et l'application documentaire étant progressivement résolues, celles qui touchaient à la cartouche de connaissance ont émergé, ce qui créa des tensions entre Temis et l'architecte innovation. Au final, celle-ci et les représentants de Temis se rejetèrent mutuellement la responsabilité de cette situation, ce qui aboutit à l'éviction de l'architecte innovation, et au « débranchement » de la brique d'extraction terminologique de Temis. La relation avec la nouvelle chef de documentation se limitant à quelques tentatives inabouties de relance de la collaboration.

D'un point de vue général, la proximité entre les chefs de projet et autres acteurs de Temis, Mondeca et Xyleme avec l'architecte innovation fut limitée par le manque de compréhension de cette dernière quant au projet en général, et par son manque de compétence dans les domaines de la représentation des connaissances et de l'extraction terminologique, en particulier. Ce déficit a été initialement sous-estimé par les membres du consortium, ce qui a été source de malentendus, et de dysfonctionnements. Une des raisons qui expliquent ces malentendus, est que l'architecte innovation, étant à l'origine du projet, a été perçue comme naturellement compétente. L'atmosphère de la relation avec la chef de documentation s'est donc détériorée, principalement avec Temis, à la fin du projet. Le manque de compétences explique aussi certainement que l'architecte innovation n'ait pas réellement été en mesure de superviser le projet, laissant une forte autonomie à chaque membre du consortium, autonomie caractéristique de l'interaction entre les membres du consortium et les deux responsables du projet chez PressPro (architecte innovation et responsable intégration informatique). L'architecte innovation n'aura finalement exercé son pouvoir vis-à-vis des membres du consortium qu'au début du projet, en imposant Xyleme comme troisième membre du consortium. Ensuite, elle n'a plus été en mesure de l'exercer, placée en position de faiblesse du fait de son manque de compréhension du projet, mais aussi, en interne, du fait du comportement très réticent des utilisateurs classiques face à l'innovation.

En synthèse, nous observons que les enjeux étaient forts de part et d'autre : refonte des tâches du service documentation de PressPro pour l'architecte innovation, réalisation d'un premier projet couplant IDE/ITM, qui plus est dans un domaine qui constitue un nouveau champ d'application pour l'extraction terminologique. La motivation qui en a découlé, a été tout aussi élevée, chez des acteurs parfaitement conscients de ces enjeux. Mais, les compétences étaient déséquilibrées, ce

qui a entraîné, nous venons de le voir, une compréhension contrastée du projet et de son évolution, et une redistribution des tâches. Le pouvoir au sein de la relation est ainsi passé de l'architecte innovation de PressPro, aux chefs de projet des entreprises du consortium lors du transfert de la réalisation de l'application documentaire à Xyleme, et de l'arrêt des *user groups*, suivi du transfert de deux utilisateurs de PressPro chez Temis. Il est revenu en fin de projet au « comité qualité » de PressPro en charge d'effectuer la « recette » du projet.

### 3.3.2.2 Interaction entre membres du consortium et utilisateurs classiques de PressPro :

Nous retrouvons ici quatre phases distinctes, dont les trois premières correspondent aux trois phases de la relation des membres du consortium avec l'architecte innovation :

1. **La première phase** correspond au lancement des ateliers utilisateurs de mai à décembre 2004, pendant lesquels les interactions concernaient la majorité des utilisateurs classiques et chaque chef de projet du consortium, ainsi que les chefs de projet ontologie et cartouche de connaissance. Le sentiment des documentalistes a alors été marqué par une certaine méfiance, mais sans franche hostilité. L'argument majeur avancé par le management de PressPro, à savoir l'enrichissement des tâches des documentalistes grâce à l'innovation, portait dans une certaine mesure. Les chefs de projet du consortium ont été incités à la prudence par l'architecte innovation de PressPro, qui craignait un rejet « violent » des documentalistes. Les interactions ont été régulières et fréquentes (hebdomadaires ou bihebdomadaires), chaque chef de projet participant à plusieurs des huit groupes de travail. Pour les chefs de projet du consortium, l'utilité a surtout été au cours de cette phase, de pouvoir mieux comprendre les habitudes de travail des documentalistes. Elle a aussi été de faire émerger quelques utilisateurs classiques particulièrement intéressés par cette innovation.
2. **La deuxième phase**, de janvier à juin 2005, correspond à la période où les ateliers utilisateurs ont cessé de fonctionner, et où deux utilisateurs classiques ont participé à la collaboration avec le consortium, dans les locaux de Temis. Cette fois l'atmosphère de l'interaction fut bien différente, les deux utilisateurs classiques étant en quelque sorte considérés comme des employés du consortium travaillant exclusivement à la réalisation de la solution innovante, avec les chefs de projet de Xyleme et de Mondeca principalement. Leurs compétences initiales en *text mining* et représentation des

connaissances étaient aussi faibles que celles de leurs collègues, mais ce sont développées grâce à ce travail.

3. **La troisième phase**, à partir de septembre 2005 à février 2006, correspond à la phase de test précédant la signature de la recette par PressPro. Une équipe de quatre utilisateurs classiques, a ainsi testé la solution avec l'architecte innovation. Ces remarques concernaient en premier lieu l'application documentaire et la brique de Mondeca. L'atmosphère de la relation fut relativement sereine, les utilisateurs classiques formulant des requêtes qui purent être globalement prises en compte par Mondeca et Xyleme, ce qui mit fin aux interactions entre ces deux éditeurs de logiciel et ces utilisateurs classiques.
4. **La quatrième phase** de mars 2006 à novembre 2006, correspond à la mise en production du système, et a vu dans un premier temps les utilisateurs classiques rejeter fermement la brique logicielle d'extraction terminologique. L'architecte innovation a quitté ses fonctions en mars 2006, alors qu'elle était l'interlocutrice presque unique de Temis. Sa remplaçante s'est alors faite la porte-parole des utilisateurs classiques en affirmant que la brique d'extraction était un « mystère » et que les résultats d'extraction et de catégorisation étaient inacceptables. Le projet avait déjà dérivé dans le temps, il s'agissait d'améliorer rapidement cette brique, ce qui n'a pas été possible, Temis découvrant alors seulement l'écart entre ce qu'attendait PressPro, et les performances de sa cartouche de connaissance, et donc l'étendue du travail restant à effectuer. Une lettre des utilisateurs classiques exigeant le débranchement de la brique logicielle de Temis, a alors conduit à l'arrêt de toute interaction entre Temis et les utilisateurs classiques.

On constate que l'atmosphère de l'interaction entre les membres du consortium et les utilisateurs classiques a été dès le départ marquée par la méfiance, les réactions de ces derniers étant redoutées par les membres du consortium. La volonté de les associer dans leur ensemble, au projet, dans le cadre d'ateliers, résultait à la fois de la nécessité de recueillir les informations relatives à leur façon de travailler, afin de définir les interfaces utilisateur, et du désir de les convaincre de l'intérêt de la future solution. Les utilisateurs classiques possédaient à la fois très peu de compétences dans les technologies employées par l'innovation ce qui a accru la distance avec les acteurs du projet chez Temis et Mondeca, une motivation limitée qui s'est transformée en hostilité pour la plupart d'entre elles, et le pouvoir d'entraver le projet du fait de leur forte syndicalisation. Il est vrai que l'enjeu du projet s'il était officiellement d'enrichir les tâches effectuées par les documentalistes, était aussi, implicitement, de réduire leur nombre.

### 3.3.2.3 Interaction des membres du consortium avec les informaticiens de PressPro :

Cette interaction s'est déroulée en trois phases :

1. **Première phase** : deux informaticiens de PressPro ; sous l'autorité du chef de projet intégration informatique, ont tenté de développer l'application documentaire, après avoir reçu une formation d'une journée sur les logiciels de Temis, Mondeca et Xyleme. Leur perception de la solution à mettre en place était beaucoup plus neutre que celle des utilisateurs, l'enjeu pour eux n'étant pas le même. Il s'agissait d'un projet informatique comme un autre, si ce n'est qu'ils ne connaissaient pas les technologies nouvelles à employer. A la surprise des membres du consortium, ils n'ont pu s'approprier cette tâche.
2. **Deuxième phase** : le chef de projet de Xyleme, avec l'appui du chef de projet de Mondeca, a spécifié l'application, les deux informaticiens travaillant sous les ordres du chef de projet de Xyleme sur cette tâche. Là encore, s'il y a eu inversion du pouvoir dans la relation étudiée, l'atmosphère de l'interaction est restée assez neutre. Mais le manque de compétences des deux informaticiens a conduit Xyleme à recruter un informaticien pour ce développement.
3. **Troisième phase** : les interactions ont cessé, la tâche étant complètement assumée par Xyleme.

Les informaticiens du service Documentation de PressPro ont fait preuve d'une motivation limitée, même s'ils n'ont montré aucune hostilité à l'égard du projet, tandis que leurs compétences en représentation des connaissances, et extraction terminologique, étaient très restreintes. Cette interaction a donc été caractérisée par une certaine neutralité de l'atmosphère, une neutralité de la position des informaticiens vis-à-vis du projet, un manque de compétences de ceux-ci, qui a conduit les membres du consortium à assumer entièrement la tâche de réalisation de l'application documentaire, ce qu'ils souhaitaient a priori absolument éviter. Le pouvoir au sein de cette relation est passé des informaticiens de PressPro aux membres du consortium à l'occasion du transfert de cette tâche.

### 3.3.2.4 Interaction des membres du consortium avec le management de PressPro :

La direction de PressPro cherchait à rester à la pointe de la technologie dans le domaine des systèmes d'information documentaires. Elle a recruté en 2002, une chef de documentation texte,

venue d'un grand magazine économique réputé dans ce domaine. L'enjeu pour elle était à la fois de rester à la pointe de la technologie, mais aussi d'augmenter la productivité des documentalistes grâce à l'automatisation de certaines de leurs tâches. L'interaction avec les membres du consortium s'est limitée initialement à une présentation en décembre 2003 au DSI de PressPro, qui a permis d'écarter le concurrent Ever, proposé par un des managers de PressPro. Ensuite, l'interaction a eu lieu de manière indirecte, via le document de spécifications, remis au Directeur des Systèmes d'Information de PressPro, et validé par celui-ci. Par la suite, les interactions avec la direction de PressPro, se sont limitées à la réunion fixant le chiffrage définitif du projet, réunion qui a rassemblé le directeur délégué de PressPro accompagné de l'architecte innovation, et les responsables commerciaux de Temis, Mondeca et Xyleme. Les difficultés du projet, n'ont pas conduit à de nouvelles interactions entre les membres du consortium et le management de PressPro. L'interaction avec le management de PressPro, est donc restée très limitée, à l'image de l'implication de celui-ci dans le projet.

### 3.3.3 Interaction fournisseur/client lors du projet Exinis :

#### 3.3.3.1 Interaction entre Temis, Mondeca et 4D Concept, et l'architecte innovation d'Exinis :

Les relations des chefs de projets de ces trois firmes, ainsi que des chefs de projet ontologie et cartouche de connaissance, avec l'architecte innovation furent permanentes, à la fois rythmées par des réunions régulières, et par des contacts impromptus en fonction de l'avancement du projet. L'architecte innovation possédait des compétences et une vision du projet qui dépassaient ses enjeux immédiats. Il a fait preuve d'une très grande motivation vis-à-vis de ce qu'il considérait comme « son » projet, auquel il consacrait la majeure partie de son temps. L'enjeu était donc pour lui très fort, car ce projet est devenu, sous son impulsion, le projet du groupe Exinis, dans le cadre du « projet groupe » Kepler. La vision et les compétences de l'architecte innovation lui ont permis d'impulser, de coordonner et de superviser le projet. Ses compétences transversales en *text mining* et en gestion des connaissances lui ont permis de dialoguer avec les acteurs du projet de Temis, Mondeca et 4D Concept, et d'orienter leur travail. Ces compétences ne faisaient toutefois pas de lui un expert des domaines évoqués. Elles l'ont aidé à prendre des initiatives telles que la constitution d'une base de connaissance virtuelle, et à réclamer des fonctionnalités nouvelles. L'atmosphère de l'interaction a ici été marquée par la proximité et la confiance entre acteurs, par un esprit de coopération, voire par l'admiration des chefs de projets et autres acteurs du projet chez les « prestataires » envers l'architecte innovation. C'est lui qui a impulsé et coordonné le

projet, appuyé par le chef de projet intégration informatique qui s'occupait du suivi opérationnel. L'architecte innovation d'Exinis a supervisé la relation avec les chefs de projet de Temis et de Mondeca (en collaboration avec le responsable chef de projet intégration informatique dans ce dernier cas) et avec le chef de projet de 4D Concept. Il a même travaillé directement avec ce dernier au développement de l'application documentaire. En interne, il a dirigé le comité de projet, dont faisaient partie les deux utilisateurs experts, et géré la collaboration avec les groupes d'utilisateurs. Il a été, par ailleurs, celui qui a géré l'interaction avec le management français et international d'Exinis. Il avait donc un réel pouvoir sur tous les acteurs du projet. En synthèse, l'interaction a été marquée par une grande intensité des échanges, nettement supérieure à celle observée entre les membres du consortium et l'architecte innovation de PressPro. Elle est aussi caractérisée par une plus grande richesse de ces échanges, due à l'attitude beaucoup plus active et aux initiatives de l'architecte innovation d'Exinis.

### 3.3.3.2 Interactions entre Temis, Mondeca et 4D Concept, et les utilisateurs d'Exinis :

Pendant toute la durée du projet, elle s'est déroulée en deux temps : une interaction directe et permanente avec les utilisateurs experts du comité de projet (dont l'une travaillait aussi avec la responsable de couplage de Mondeca), et une interaction indirecte avec des groupes d'utilisateurs, au travers d'ateliers organisés par ces utilisateurs experts. Les chefs de projet de Temis, Xyleme, et 4D Concept n'ont pas été en contact direct avec les utilisateurs finaux classiques.

Les interactions avec les utilisateurs experts ont été multiples en ce sens que ces utilisateurs ont travaillé à la conception de la cartouche de connaissance, à celle de l'ontologie, à celle des RAC, et à celle de l'application documentaire, directement avec les différents acteurs du projet chez Temis, Mondeca et 4D Concept. Ils ont collaboré de manière suivie, en tant que membres à part entière de l'équipe-projet. Ces deux utilisateurs dotés de compétences initiales en gestion des connaissances, n'en étaient toutefois pas des experts, mais en possédaient un premier niveau de connaissance, étaient sensibilisés à ces enjeux, et en percevaient l'intérêt dans le cadre de leur activité professionnelle. L'interaction avec les utilisateurs experts s'est donc déroulée dans une atmosphère favorable, résultat d'une forte motivation des deux parties et de la capacité à collaborer de ces deux utilisateurs. La richesse de l'interaction a aussi découlé du fait que l'interaction avait débuté très tôt, grâce au développement rapide d'un prototype qui avait permis de faire réagir les utilisateurs experts.

### 3.3.3.3 Interactions entre Temis, Mondeca et 4D Concept et les informaticiens d'Exinis :

Les informaticiens d'Exinis, dont le chef de projet intégration informatique, n'ont pas eu à effectuer de tâche relevant directement du projet. L'interaction entre les prestataires et les informaticiens d'Exinis, concerne surtout, pour ce qui est d'Exinis, les informaticiens du comité de projet, i.e. d'une part l'architecte innovation (interaction que nous avons déjà évoquée) et le chef de projet intégration informatique, spécialiste des bases de données qui a beaucoup travaillé avec chef de projet ontologie de Mondeca. Ce chef de projet intégration informatique a consacré l'essentiel de son activité à l'organisation et au suivi opérationnel des tâches du projet, supervisé par l'architecte innovation. Cela l'a amené à travailler avec les chefs de projet des prestataires. L'interaction entre le chef de projet ontologie de Mondeca et le chef de projet intégration informatique d'Exinis a été facilitée par les compétences de ce dernier en bases de données, à la différence de ce qui s'est produit au cours du projet PressPro. L'interaction présentait ici des caractéristiques proches de l'interaction avec les utilisateurs décrite plus haut, du fait du niveau de compétences et de la motivation de ce chef de projet intégration informatique, toutes deux moins élevées cependant que celles de l'architecte innovation.

### 3.3.3.4 Interactions entre Temis, Mondeca et 4D Concept et le management d'Exinis :

La première implication du management d'Exinis date du feu vert donné par le CIO (Chief Information Officer ou Directeur des Systèmes d'Information) d'Exinis France, à l'architecte innovation, pour qu'il fasse réaliser un démonstrateur de la solution qu'il a imaginée. Ensuite, l'interaction avec le management d'Exinis, s'est pratiquement toujours déroulée via l'architecte innovation. Toutefois, en octobre 2004, l'architecte innovation, accompagné des chefs de projet de Temis et de Mondeca, a fait une démonstration de la solution aux managers d'Exinis USA.

L'organisation du projet au niveau du management d'Exinis France était assez sophistiquée puisqu'elle comprenait, selon la terminologie appliquée par Exinis, le DSI (ou CIO), un « sponsor de projet », un « architecte innovation » et un « chef de projet intégration informatique ». Une telle organisation ne signifiait pas que le management soutenait aveuglément le projet, comme l'a montré sa suspension provisoire, ou le soin pris par l'architecte innovation à valoriser l'innovation aux yeux de sa hiérarchie. L'architecte innovation a donc développé une véritable stratégie pour expliquer ce projet aux managers d'Exinis. La visibilité du projet a été d'autant plus grande qu'il s'inscrivait dans un projet plus vaste, le projet Kepler. Cela a contribué à crédibiliser le

projet en interne, mais aussi fait peser des contraintes fortes sur ceux qui en avaient la charge chez Exinis.

En conclusion, l'interaction avec le management d'Exinis a donc été ponctuelle, limitée, et encadrée par l'architecte innovation.

### 3.3.3.5 Comparaison entre les projets PressPro et Exinis, du point de vue des interactions fournisseur/client :

Les interactions fournisseur/client ont été différentes d'un projet à l'autre pour des raisons qui tiennent, en particulier, aux compétences et aux motivations des acteurs impliqués chez le client. Ces interactions se distinguent ainsi nettement de par leur intensité et leur atmosphère. Une différence majeure entre les projets PressPro et Exinis vient de ce que le pouvoir est clairement revenu à l'architecte innovation d'Exinis, dans le deuxième cas, alors que dans le premier cas, le pouvoir est en partie revenu aux utilisateurs. C'est une des raisons majeures de l'échec de ce projet, étant donné l'hostilité plus ou moins ouverte qu'ils ont manifesté à l'égard du déploiement de l'innovation. A la différence du projet PressPro, lors du projet Exinis, l'architecte innovation a su développer une stratégie qui a permis de bénéficier des indispensables apports des utilisateurs classiques, tout en parvenant à encadrer leurs contributions. Il a donc su mieux interagir avec les utilisateurs classiques, que l'architecte innovation de PressPro. Il a su aussi intégrer des utilisateurs experts au projet. Il en est allé de même en ce qui concerne la relation avec son propre management, que l'architecte innovation d'Exinis a pensé et organisé.

### 3.3.4 Interaction entre Temis et Mondeca :

Cette interaction entre éditeurs de logiciel, a été permanente tout au long des projets PressPro et Mondeca. Ainsi, l'intensité de l'interaction entre Temis et Mondeca a été élevée, du fait de la nécessité de concevoir et développer une passerelle de couplage, pendant le projet PressPro. Elle a été organisée, mais sans exercice d'autorité hiérarchique, principalement par la chef de projet couplage chez Mondeca, et a donné lieu à de nombreuses interactions. La prolifération de ces interactions s'est avérée extrêmement chronophage pour chacun des acteurs concernés (chefs de projet couplage, cartouche de connaissance et ontologie), mais soumis à la nécessité d'avancer sur le projet, la prise de recul par rapport à cette question est restée limitée pendant la durée du projet PressPro, malgré les insatisfactions ressenties par ses acteurs. RAC, cartouche de connaissance et

ontologie étant interdépendantes, le problème était de savoir ce qui devait être figé et ce qui devait évoluer : pour faire évoluer certaines tâches, il fallait en figer d'autres, et pour cela il fallait soit que les acteurs se mettent d'accord, soit qu'une autorité hiérarchique s'exerçât. Or, les organisations étaient faiblement couplées, et personne n'était en mesure d'exercer une telle autorité. Il fallait donc qu'un accord intervienne entre plusieurs acteurs, et les interactions ont proliféré. Une telle prolifération pouvait s'avérer dangereuse pour le projet, enserré dans un tissu d'interactions tellement dense qu'il risquait d'en entraver la progression.

Nous analysons les caractéristiques de cette interaction en reprenant les items que nous avons retenus du modèle d'interaction (Hakansson, 1982 ; Michel, Salle et Valla, 1996) :

- L'atmosphère de la relation a été caractérisée par une volonté de coopération et une absence de conflit. Les deux sociétés sont des start-up parisiennes, créées au début des années 2000 présentant un certain nombre de similitudes tant au niveau de leur taille, que des caractéristiques de leur équipe dirigeante (cadres de l'industrie informatique, d'une même classe d'âge, de même formation) ou encore de leur organisation qui accorde une grande place à la R&D, équipe de R&D où l'on d'ailleurs retrouvait des chercheurs possédant le même type de formation (informaticiens, formés à la gestion des connaissances ou l'analyse sémantique, et ayant travaillé dans la recherche). La distance sociale et culturelle était en conséquence faible. Le sentiment qu'elles partageaient était celui de la complémentarité de leurs technologies respectives, et leur volonté commune de proposer sur le marché une offre couplée, la première du genre, était forte, dans la perspective de tirer avantage d'une position de « premier entrant ». L'atmosphère de la relation était donc très favorable au bon déroulement des interactions entre les deux firmes, dans un contexte d'équilibre de pouvoir entre les deux firmes. On note à l'inverse que Temis et Mondeca ont expliqué leur méfiance vis-à-vis de Xyleme avant le projet PressPro, par le fait qu'il s'agissait d'une start-up ayant levé des fonds importants auprès de capitaux-risqueurs, et n'ayant jamais eu réellement de clients.
- L'environnement général était aussi favorable à l'interaction entre les deux éditeurs, puisqu'il voyait la maturation d'offres diverses dans le domaine de la représentation des connaissances et du *text mining*, dans un contexte de développement du Web sémantique. Une telle maturation fournissait un cadre favorable à la convergence entre ces différentes applications. Elle poussait ainsi les éditeurs de logiciel concernés à travailler à la réalisation d'une plate-forme logicielle intégrant ces différentes applications, dans le but

de figurer parmi les premiers acteurs à proposer une plate-forme logicielle combinant extraction d'information et système de gestion des connaissances.

Les caractéristiques des interactions entre Temis et Mondeca, sont par ailleurs proches de celles, sur lesquelles nous ne revenons pas en détail, qui se sont déroulées entre ces deux éditeurs de logiciel et Xyleme, puis entre ces deux éditeurs de logiciel et 4D Concept. D'un point de vue général, cela découle, à notre sens, de la proximité qui existe entre ces différentes organisations du fait leur culture technologique, de leur taille, et de leur motivation quant à la réussite du ou des projets auxquelles elles ont participé.

Enfin, la collaboration entre Temis et Mondeca, a conduit les acteurs de ces deux éditeurs de logiciel, engagés dans les projets PressPro et Exinis, à penser qu'une association étroite, voire une fusion entre les deux entreprises serait souhaitable. Toutefois, une telle fusion semblait peu probable à court terme, les deux éditeurs souhaitant rester focalisés sur leur technologie propre.

### *3.3.5 Conclusion sur la nature des interactions entre acteurs :*

En conclusion, nous soulignons le caractère multidimensionnel des interactions fournisseur/client dans le cadre des projets étudiés. Ces interactions ont été à la fois diverses dans le cadre d'un même projet, et différentes d'un projet à l'autre. Cela atteste de la multiplicité des rôles et des acteurs impliqués dans les projets. L'analyse de ces interactions éclaire la nature des rôles, ainsi que celle des contributions de ces acteurs. Cela nous conduit à analyser maintenant la façon avec laquelle ces rôles ont émergé, et plus globalement comment les projets se sont progressivement structurés. Cela nous conduira ensuite (point 4) à discuter de la contribution des clients au processus d'innovation.

## **3.4 Une structuration de projet émergente :**

### *3.4.1 Introduction :*

Nous nous intéressons maintenant à la structuration progressive des projets de création automatisée de base de connaissance. L'analyse de cette structuration permet notamment d'éclairer les modalités de coordination entre les organisations qui ont participé à ces projets, et de préciser le rôle et les caractéristiques de différents acteurs.

Le besoin de structuration des projets étudiés vient de ce qu'il requiert l'accomplissement de tâches qui sont à la fois nouvelles, interdépendantes et réalisées par des organisations faiblement couplées. Ces tâches sont de deux ordres. Tout d'abord, deux d'entre elles consistent à appliquer une technologie (l'extraction d'information et la représentation des connaissances) à un domaine nouveau (articles de la presse « people » ou textes de jurisprudence). Les trois autres sont entièrement nouvelles, le couplage IDE/ITM, la définition des règles d'acquisition de connaissance, mais aussi le développement de l'interface utilisateur (l'application documentaire) de ce nouveau type de solution.

Nous observons que la responsabilité de chacune de ces tâches était a priori attribuée à un acteur précis, à l'exception de l'application documentaire, qui a été assumée par différents acteurs au fil des projets. La caractérisation du type d'acteur susceptible d'endosser cette tâche fait donc aussi partie des éléments de structuration qui ont émergé. Par ailleurs, les projets ont permis de caractériser les types de collaboration à mettre en œuvre par les responsables de chacune des tâches, et de répondre aux questions suivantes : Avec qui collaborer ? Selon quelles modalités ?

Nous estimons que les collaborations interorganisationnelles ont été déterminantes quant à la structuration des projets étudiés, c'est pourquoi nous ne revenons pas en détail, ici, sur les collaborations intra-organisationnelles observées. Ainsi, le développement d'une cartouche de connaissance juridique dans le cadre du projet PressPro, malgré la nouveauté du domaine, correspondait à une tâche habituelle pour Temis. Mais, comme ce travail s'inscrivait dans un projet de création de base de connaissance, il a dû être réalisé de manière concurrente et concertée avec la conception de l'ontologie et des RAC par Mondeca, ce qui était, cette fois, entièrement nouveau pour Temis. Il en est allé de même pour la création d'une ontologie par Mondeca, en collaboration avec Temis. La conception de la passerelle OntoPop et la définition de RAC personnalisées, ont aussi nécessité une collaboration étroite entre Temis et Mondeca, même si Mondeca en a assumé la plus grande part. Enfin, l'acteur chargé de la réalisation de l'application documentaire a dû collaborer avec le chef de projet ontologie de Mondeca. De plus, ces différentes tâches ont dû être coordonnées, ce qui pose ensuite la question de la coordination globale et du pilotage du projet. En synthèse, la structuration de projet passe par la définition de différentes modalités : les modalités de collaboration interorganisationnelle relative à une tâche donnée, les modalités de coordination des tâches entre elles, et les modalités de coordination globale de projet. C'est ce que nous nous proposons d'examiner maintenant.

### 3.4.2 Structuration du projet PressPro :

Au départ, l'architecte innovation de PressPro avait envisagé une structuration de projet en deux parties. D'une part, le développement de l'application documentaire confiée aux informaticiens de PressPro. D'autre part, l'intégration des briques logicielles de Temis, Mondeca et Xyleme, placées sous la responsabilité de Xyleme. Cette intégration englobait le couplage IDE/ITM, placé sous la responsabilité de la chef de projet couplage de Mondeca. C'est ainsi que le projet a été structuré, jusqu'au transfert de la responsabilité de l'application documentaire de PressPro à Xyleme. Un deuxième type de structuration a ensuite été observé. Nous revenons maintenant sur ces deux phases.

#### **Première phase de structuration du projet PressPro :**

Dès le début du projet PressPro, Mondeca a recruté une chef de projet couplage. Lors du projet PressPro, celle-ci a conçu les logiciels de la passerelle OntoPop en collaboration avec l'équipe-produit de Mondeca. Ce travail a été effectué en procédant à des consultations régulières du chef de projet cartouche de connaissance de Temis. La chef de projet couplage a aussi coordonné ses travaux portant sur la définition de RAC personnalisées, avec ceux du chef de projet cartouche de connaissance de Temis et avec ceux du chef de projet Ontologie de Mondeca. Cette coordination liée au couplage moteur d'extraction/système de représentation des connaissances, a nécessité un travail intense, mais n'a pas posé de problèmes techniques insurmontables. Elle a bénéficié de la volonté de Temis et Mondeca de voir aboutir le processus d'innovation, dans une perspective de constitution de plate-forme applicative, ainsi que de la proximité (intellectuelle et culturelle) entre ces trois protagonistes, que nous avons évoquées au point 3.3.4 à propos des interactions entre fournisseurs. Cette coordination s'est donc déroulée de la manière prévue, mais a posé problème du fait du foisonnement des interactions entre acteurs qu'elle a généré. Un tel foisonnement s'est révélé lourd à gérer et insatisfaisant pour les acteurs concernés, à tel point qu'à la fin du projet PressPro, il est apparu indispensable de définir une ingénierie précise, la démarche OntoPop. Le projet PressPro a donc permis à Temis et à Mondeca d'explorer les modalités du couplage entre un moteur d'extraction d'information et un système de gestion de base de connaissance. Ce couplage a requis une coordination forte qui a été assumée avec succès par la chef de projet couplage.

En ce qui concerne l'application documentaire, PressPro avait choisi d'en prendre la responsabilité mais a échoué. Cet échec tient à notre sens au manque de compréhension globale du projet et des logiciels qui le fondaient, ainsi qu'au manque de compétences des informaticiens de PressPro dans le domaine du développement d'interfaces utilisateur. Ces informaticiens avaient été formés pendant quelques jours, au logiciel IDE par Temis, à la base de connaissance ITM par Mondeca, et au logiciel XML Server par Xyleme, au début du projet. A la suite de cela, ils devaient travailler avec les futurs utilisateurs du système à spécifier l'application. Aux yeux des membres du consortium, la courte formation prodiguée était suffisante pour les rendre aptes à développer l'application documentaire. Mais, en fait, elle n'a pas permis aux informaticiens de PressPro d'acquérir un niveau d'expertise suffisant en *text mining* et en représentation des connaissances pour définir, avec les utilisateurs, l'interface d'un système qui repose sur ces technologies. Cela démontre, à notre sens, l'écart de compétences entre le client et les fournisseurs, écart dont ni l'un ni les autres n'avaient perçu l'ampleur. Focalisé sur sa technologie et souhaitant ne pas endosser cette tâche, chaque éditeur de logiciel a semblé en sous-estimer la difficulté pour les informaticiens de PressPro.

En ce qui concerne la coordination globale du travail des trois éditeurs du consortium, le chef de projet de Xyleme, qui en avait initialement la responsabilité, n'a pu l'assumer. Cela vient du fait que les différentes collaborations entre Temis, Xyleme et Mondeca, n'ont pas requis le même niveau de coordination. En effet, la collaboration entre Temis et Xyleme s'est limitée au transfert de la base d'articles de PressPro au format XML, tandis que la collaboration entre Mondeca et Xyleme a consisté à définir une passerelle entre les logiciels ITM et XML Server. Les collaborations de Xyleme avec Mondeca et Temis, ont été limitées parce que le logiciel de Xyleme, XML Server, constituait un module de la solution, et qu'à ce titre, sa connexion au système de création de base de connaissance ne nécessitait pas de travail particulier, une fois récupérés les articles stockés dans l'ancienne base de données de PressPro, et une fois réglées les questions de format de données. A l'inverse la collaboration entre Temis et Mondeca à propos des RAC, a été intense comme nous l'avons vu, et a été prise en charge par la chef de projet couplage. Enfin, en ce qui concerne la collaboration tripartite, entre Xyleme, Mondeca et Temis, si une première coordination globale a été nécessaire au moment de la contractualisation, nous avons pu constater qu'ensuite les collaborations ne se sont faites que deux à deux. Ainsi, la définition d'un « chef de projet logiciel » en charge de la coordination de l'action des trois membres du consortium s'est avérée inopérante dans le cadre du projet PressPro.

Enfin, lors de cette première phase, la coordination globale du projet dans son ensemble, est revenue implicitement à l'architecte innovation de PressPro. Cette coordination n'avait pas été pensée en tant que telle par l'architecte innovation, qui n'en avait pas perçu la nécessité. Elle est donc restée légère, ce dont les membres du consortium se sont plaints. Ils auraient souhaité s'appuyer sur un acteur proposant des orientations et suggérant des choix clairs, dès lors que survenaient les difficultés inhérentes au développement et au déploiement d'une innovation logicielle.

### **Deuxième phase de structuration du projet PressPro :**

A la suite de l'échec de PressPro, l'application documentaire a été attribuée Xyleme, qui ne possédait pourtant pas d'expérience en développement d'interface utilisateur. Xyleme a su développer cette application documentaire, ce qui nécessitait tout d'abord d'interagir avec les utilisateurs. Ensuite, cela demandait de collaborer avec le chef de projet ontologie, pour que cette application corresponde à l'ontologie définie. Il y avait donc là, un besoin de coordination, coordination qui a été réalisée en binôme par le chef de projet de Xyleme et le chef de projet ontologie de Mondeca. La proximité culturelle et technologique entre Xyleme et Mondeca (cf. 3.3.4 de ce chapitre), a facilité ce travail, tout comme leur forte motivation à faire aboutir le projet. Cette motivation partagée et le constat du manque de compétences tant chez le client que dans les sociétés de services informatiques, ont donc poussé Xyleme à accomplir une tâche inhabituelle, pour laquelle un informaticien, dédié au développement de l'application spécifiée par le chef de projet, a été recruté.

Par ailleurs, lors de cette deuxième phase, la coordination globale du projet PressPro, par l'architecte innovation, est restée faible. C'est, sans doute, l'une des causes de l'échec du développement de la cartouche de connaissance « people ». En effet, la tâche de développement de cette cartouche a été effectuée par Temis au travers d'une collaboration ponctuelle et exclusive avec l'architecte innovation. Cette dernière n'a alors pas fait remonter d'information alarmante à ce sujet. Pendant ce temps, les utilisateurs ont travaillé à l'amélioration des autres briques logicielles du système. Une fois celles-ci au point, le projet avait accumulé un tel retard qu'il devenait obligatoire de finaliser la cartouche de connaissance « people » au plus vite. Les problèmes rencontrés par les documentalistes avec la première version fournie par Temis n'ont pas pu être résolus rapidement, ce qui a débouché sur le rejet définitif de cette cartouche. Une meilleure coordination globale du projet aurait pu permettre d'éviter cet échec.

En conclusion, le projet PressPro a fait apparaître la nécessité de prendre en compte trois types de structuration : une coordination liée au couplage extraction terminologique/représentation des connaissances, une coordination liée à la cohérence entre l'application documentaire et l'ontologie, et une coordination globale du projet. En revanche, il a montré qu'une coordination tripartite entre membres du consortium était inutile, du fait de la modularité du composant de Xyleme.

### 3.4.3 Structuration du projet Exinis :

Nous allons maintenant présenter la façon avec laquelle ces trois types de structuration, se sont mises en place au cours du projet Exinis. Contrairement au projet PressPro, on ne peut pas distinguer de phases très distinctes du point de vue de la structuration de ce projet, qui est donc marquée par une certaine linéarité.

#### **Coordination nécessaire au couplage IDE/ITM :**

Par rapport au projet PressPro, la mise en œuvre de la méthodologie OntoPop, notamment le découpage du processus en cinq phases, a facilité la définition des RAC. Cette mise en œuvre a conduit à une organisation plus efficace des interactions entre acteurs, et s'est traduite par une baisse du nombre de ces interactions. Ainsi, la méthodologie a été appliquée, dans une large mesure, et a donné satisfaction aux acteurs du projet. Le non respect de certains aspects de la méthodologie, est venu du manque de temps ou de ressources humaines, que Mondeca, Temis ou Exinis ont pu y consacrer.

#### **Coordination nécessaire à la cohérence entre l'application documentaire et l'ontologie :**

4D Concept a été chargé du développement de l'application documentaire. Le chef de projet de 4D Concept a pu accomplir cette tâche parce qu'il possédait des compétences en développement d'interfaces utilisateur, parce qu'il connaissait le métier de rédacteur juridique chez Exinis, ce qui a facilité son travail avec les utilisateurs, et parce qu'il était fortement motivé. Cette motivation résultait du désir de devenir *VAR* (value added reseller, i.e. revendeur « officiel ») de la solution couplée IDE/ITM, en capitalisant sur les compétences acquises lors de ce projet. La coordination entre la réalisation de l'application documentaire et celle de l'ontologie a été, quant à elle, le fruit de la collaboration de deux acteurs, le chef de projet de 4D Concept et le chef de

projet ontologie de Mondeca. Ces deux acteurs ont bénéficié des apports du chef de projet intégration informatique d'Exinis. Celui-ci a apporté une contribution précieuse au travail de ces deux chefs de projet de par sa connaissance des bases de données en général, et de celles d'Exinis en particulier.

### **Coordination globale et pilotage du projet Exinis :**

Les personnes impliquées dans le projet chez Exinis ont été plus nombreuses que chez PressPro, et leurs différents rôles ont été définis dès le départ (nous indiquons entre guillemets les dénominations utilisées par Exinis) : un « sponsor » (directeur des équipes éditoriales, qui rassemblent les utilisateurs classiques), le DSI, un « comité de projet ». Ce comité était composé de l' « architecte innovation » (terminologie que nous avons retenue), du « chef de projet intégration informatique » spécialiste des bases de données, de deux rédactrices juridiques possédant une formation en gestion des connaissances (que nous nommons utilisateurs experts), d'un informaticien en charge de la qualité, et des groupes d'utilisateurs classiques amenés à participer à des ateliers. Il est vrai que ce projet s'inscrivait dans le cadre du projet Kepler, qui constituait un enjeu de taille pour le groupe Exinis.

Parmi les membres du comité de projet, l'interaction avec Temis, Mondeca et 4D Concept a été en grande partie prise en charge par l'architecte innovation, assisté, au niveau de la gestion opérationnelle du projet, du chef de projet intégration informatique. L'architecte innovation a été reconnu par Temis, Mondeca et 4D Concept comme le seul acteur, chez Exinis, à en posséder une compréhension précise et à en fixer des orientations claires. Le chef de projet intégration informatique, quant à lui, s'est surtout occupé du suivi quotidien du projet (définition et attribution des tâches, suivi du planning), sous la supervision de l'architecte innovation. La coordination globale du projet a donc été assurée par l'architecte innovation. Concrètement, son rôle a consisté à assurer un suivi global des tâches, à faire remonter des problèmes majeurs et à s'assurer de leur traitement par les acteurs concernés : il en a été ainsi pour les difficultés de couplage IDE/ITM, et pour les temps d'accès à la base ITM. Il a aussi veillé à impliquer de manière régulière les utilisateurs classiques, tout en fixant un cadre stricte à leur participation. L'architecte innovation s'est aussi efforcé de coordonner la progression des tâches à effectuer dans le cadre du projet avec la nécessité de montrer, à certains moments, au management général d'Exinis des applications concrètes et édifiantes de l'innovation. C'est ainsi que l'architecte innovation a pris des initiatives, et fait réaliser quelques développements informatiques,

relativement mineurs, alors qu'ils ne s'imposaient pas dans le cadre du projet. L'architecte innovation d'Exinis a donc agi en véritable pilote du projet, et ne s'est pas contenté de coordonner les différentes tâches.

Nous mettons ainsi en évidence le fait que l'architecte innovation d'Exinis a joué un rôle, qui correspond à celui que la littérature en management de projet (Clark et Wheelwright, 1992) appelle « heavy weight project manager » (Clark et Fujimoto, 1991), ou encore « directeur de projet » (Giard et Midler, 1993 ; Midler, 1993). Il apparaît que ce rôle de « directeur de projet » a été en grande partie assumé, dans le cadre du projet Exinis, par l'architecte innovation qui disposait d'une large délégation de la part de la direction générale, et possédait un poids beaucoup plus élevé que le chef de projet intégration informatique. Le rôle de ce dernier se rapproche, quant à lui, de celui que la littérature qualifie de « light weight product manager » (Clark et Wheelwright, 1992), qui effectue la planification et le suivi quotidien des tâches à réaliser dans le cadre d'un projet. Or le modèle d'organisation de la conception appelé « heavy weight project management » centrée autour de la notion de « heavy weight project manager » ne prend pas en compte la notion de « light weight product manager ». C'est pourquoi, considérant que l'organisation du projet Exinis correspond au modèle « heavy weight project management », nous choisissons de qualifier la fonction du chef de projet intégration informatique, dans le cadre de ce projet, de « gestionnaire de projet ».

La prise en compte de la littérature relative au management de projet de systèmes d'information (Marciniak et Rowe, 1997, Morley, Hughes, Leblanc, Hughes, 2005 ; Morley, 2008) nous conduit à tenter de définir ces rôles par rapport aux notions de maître d'ouvrage et de maître d'œuvre, bien que celles-ci ne soient pas normalisées dans le domaine des systèmes d'information, et qu'elles soient pratiquement absentes de la littérature non francophone. Selon la littérature francophone en systèmes d'information, qui recoupe en cela la littérature relative au management de projet industriel (Charue-Duboc, 1997), le maître d'ouvrage est l'acteur qui est responsable de la définition des besoins, qui engage les fonds et qui sera propriétaire de l'« ouvrage » une fois celui-ci terminé. Il s'agit dans le cas de projets de système d'information, généralement d'une direction « métier ». La traduction anglo-saxonne de ce terme, est variable (« client », « project owner », voire « sponsor »), mais elle contribue à éclairer cette notion. Pour ce qui est de la notion de « maître d'œuvre », cette même littérature renvoie au rôle de l'acteur responsable de la réalisation de l'ouvrage, qui met en œuvre les moyens techniques nécessaires à la satisfaction du besoin préalablement défini. La traduction anglaise de ce terme s'avère elle aussi éclairante :

« project supervisor », « project manager » ou encore « lead contractor ». Dans le cas des systèmes d'information, le rôle de maître d'œuvre est souvent joué par la DSI de l'entreprise.

Dans quelle mesure de tels rôles ont-ils été définis et assumés dans le cadre des projets PressPro et Exinis ? Pour PressPro, il apparaît que l'organisation du projet, dont la responsabilité reposait en interne sur le tandem architecte innovation/chef de projet intégration informatique, s'appuyait clairement sur ces deux notions : l'architecte innovation de PressPro étant le maître d'ouvrage, et le chef de projet intégration informatique le maître d'œuvre. A l'inverse, chez Exinis, où de nombreux rôles ont été attribués, les choses étaient a priori moins claires : le maître d'ouvrage en tant que propriétaire du produit final aurait pu être le directeur des équipes éditoriales appelé « sponsor » en interne, tandis que le maître d'œuvre aurait pu être le DSI. En fait, leur rôle a été réel, mais limité et ponctuel. L'architecte innovation d'Exinis semble avoir joué à la fois une grande partie du rôle de maître d'ouvrage (au travers de la définition des besoins, notamment) et une partie du rôle de maître d'œuvre (via, par exemple, la gestion d'une équipe d'informaticiens d'Exinis travaillant sur le projet). En ce qui le concerne, le chef de projet intégration informatique, ou comme avons choisi de l'appeler, le gestionnaire de projet, a joué une grande partie du rôle de maître d'œuvre.

Notre conclusion sur ce point, est que, dans le cas d'une innovation présentant un caractère exploratoire, la distinction formelle entre maîtrise d'ouvrage et maîtrise d'œuvre n'est pas forcément possible ni souhaitable (projet PressPro), alors qu'une fusion partielle de ces rôles, telle qu'observée dans le projet Exinis, apparaît plus adaptée. Ce caractère exploratoire est illustré, en particulier, par la relative imprécision de la définition initiale des spécifications. Cette imprécision a été constatée dans les deux projets étudiés, et spécialement dans le projet PressPro. Elle était d'autant plus grande que l'application était radicalement innovante, ce qui rendait contreproductif le découpage classique entre maître d'ouvrage et maître d'œuvre. Nous en voulons pour preuve les différents contrats additionnels qui ont été signés au cours des deux projets, concernant des tâches qui avaient été mal spécifiées, ou qui n'avaient pas été prises en compte dans les spécifications initiales du projet. Dans ces conditions, on comprend que le travail des fournisseurs avec un acteur « directeur de projet », qui associe une grande partie des fonctions de maître d'ouvrage et de maître d'œuvre, facilite et accélère la réorientation du projet, en comparaison avec une situation où ces mêmes fournisseurs travaillent avec un maître d'œuvre qui n'a pas vocation à redéfinir les spécifications rédigées par le maître d'ouvrage.

### 3.4.4 Une structuration définie après une phase d'exploration :

Les deux projets étudiés ont permis de mettre en évidence trois niveaux de structuration qui sont, in fine, la coordination globale du projet, le couplage IDE/ITM et la coordination application documentaire/ontologie. Nous mettons en évidence les rôles des acteurs qui participent à cette structuration, et leurs caractéristiques propres, notamment leurs compétences.

#### 1. La coordination globale du projet :

La coordination globale, et au-delà, le pilotage du projet, par un acteur assumant le rôle de «heavy weight project manager», ou de « directeur de projet », nous paraît indispensable pour mener à bien un projet innovant porté par des petites structures faiblement couplées. Cet acteur doit assurer la coordination globale des tâches qu'il englobe, et, idéalement, l'orienter et gérer ses inévitables inflexions. Pour cela, il doit interagir régulièrement avec chacun des acteurs du projet, ainsi qu'avec le management de sa propre entreprise. Il peut être assisté par un gestionnaire de projet, qui en assume le suivi opérationnel et avec lequel il collabore étroitement.

Cette tâche semble devoir être, dans le cas d'une innovation radicale multi-acteurs, assumée par le client. En effet, un acteur situé chez le client semble dans une situation favorable pour, à la fois, appréhender les enjeux du projet pour le client, préparer les actions visant à convaincre le management, et organiser l'apport des utilisateurs. Un acteur situé chez le client nous semble par ailleurs idéalement placé pour faire le lien entre les différents fournisseurs avec lesquels il a contractualisé. Dans ces conditions, ce rôle revient logiquement à l'architecte innovation du client, dans la mesure où celui-ci possède une compréhension précise du projet et de ses enjeux, et dans la mesure où il est doté de compétences transversales relatives aux technologies employées, lui permettant d'imaginer ce qu'elles peuvent apporter à son entreprise et lui permettant de collaborer avec les éditeurs de logiciel concernés.

#### 2. Le couplage IDE/ITM :

La méthodologie OntoPop permet de structurer la démarche visant à définir les règles d'acquisition de connaissance, et donc le couplage entre IDE et ITM. Elle structure fortement le travail de l'ontologue, du linguiste, et du responsable de couplage. La responsabilité du couplage entre logiciels peut difficilement revenir au client, du fait de son manque de compétence, car elle

repose sur un haut niveau d'expertise dans les technologies employées par chacun des logiciels couplés. Elle repose même sur la construction de compétences fondées sur les compétences relatives à ces technologies nouvelles. C'est pourquoi, de telles compétences étant rares, cette responsabilité semble ne pouvoir être attribuée qu'à un acteur situé chez l'un des deux éditeurs de logiciel concernés, possédant des compétences en gestion des connaissances et en *text mining*.

### **3. La cohérence entre l'application documentaire et l'ontologie :**

Le besoin de coordination entre le développement de l'application documentaire et celui de l'ontologie est réel, mais il est moins poussé que celui requis par le couplage IDE/ITM. Cette coordination n'a pas donné lieu à la définition d'une méthodologie précise. Elle revient naturellement aux acteurs qui ont la charge de l'ontologie et de l'application documentaire, la conception de l'ontologie revenant par définition au chef de projet ontologie. L'acteur en charge de l'application documentaire doit, quant à lui, posséder des compétences relatives au développement d'interfaces utilisateur, ce qui est le cas des SSII. Il doit, aussi, idéalement, posséder des compétences relatives aux usages et être capable d'interagir avec les utilisateurs. Bien sûr, il doit, toujours dans l'idéal, posséder une certaine compréhension des systèmes de gestion des connaissances, pour travailler à la cohérence entre l'ontologie et l'application documentaire. Une telle combinaison de compétences étant rares, le rôle peut être endossé par un acteur possédant une partie d'entre elles et souhaitant acquérir celles qui lui font défaut.

#### *3.4.5 La question de l'encastrement de la coordination :*

Ainsi, le processus de modularisation partant de composants indépendants, « quasi-modulaires », que nous avons étudié, passe par la définition d'une structuration de projet, qui englobe d'une part une coordination globale par un « directeur de projet », et d'autre part certaines coordinations interorganisationnelles. Finalement, une fois l'architecture produite et la méthodologie liée à la mise en œuvre définies, le besoin de certains types de coordination interorganisationnelle subsiste donc, surtout parce que Temis et Mondeca restent des entreprises indépendantes proposant des produits interdépendants. Cette indépendance est susceptible de perdurer dans la mesure où les deux sociétés concentrent leurs efforts sur leurs produits et technologies propres, et dans la mesure où la solution couplée ne constitue pas le cœur de leur activité. Tant que ces logiciels seront portés par deux organisations distinctes, l'encastrement de la coordination (Sanchez et Mahoney, 1996) dans la structure quasi-modulaire de l'innovation ne

sera donc pas réalisé. Toutefois, il n'est pas impossible qu'à l'avenir Temis et Mondeca fusionnent, ou que l'un de ces éditeurs développe un produit équivalent à celui de son partenaire (i.e., un système de représentation des connaissances pour Temis, ou un moteur d'extraction terminologique pour Mondeca). Si tel devait être le cas, alors le système de création automatisée de base de connaissance serait complètement modularisé, et l'affirmation selon laquelle « les produits conçoivent des organisations » serait vérifiée (Sanchez et Mahoney, 1996).

En conclusion, l'expérience acquise lors des deux premiers projets a permis de faire émerger les enjeux de la structuration de projet de « création de base de connaissance », dans un contexte de couplage faible entre organisations. Il apparaît que cette structuration est un des enjeux clefs de ces projets, donc du processus d'innovation. Les enseignements relatifs à cette structuration constituent aussi un des éléments majeurs de capitalisation des deux projets étudiés. Ils soulignent les compétences des acteurs, ainsi que certains aspects du rôle des clients dans les projets. Nous revenons maintenant plus en détail sur ce dernier point.

#### **4. LES ROLES JOUES PAR LE CLIENT DANS LE PROCESSUS D'INNOVATION : ENRICHISSEMENT DE LA NOTION DE *LEAD USER*.**

##### **4.1 Introduction :**

Nous avons déjà évoqué le rôle du client dans le processus d'innovation étudié, au cours des différentes parties qui précèdent. Nous revenons maintenant sur ce point, afin de déterminer la réalité de ce rôle, ainsi que ces éventuelles limites, et discuter la littérature relative au concept de *lead user*. L'analyse des projets PressPro et Exinis, nous permet de souligner le rôle du client dans l'émergence de l'innovation. Il s'est manifesté à trois moments : à la structuration du projet, dans le cours du développement et en fin de projet pour son déploiement. Nous distinguons principalement trois niveaux chez le client qui ont contribué à la structuration de l'offre : l'acteur qui prend l'initiative de lancer le projet, les utilisateurs du logiciel et les informaticiens.

## 4.2 Le rôle du client à plusieurs niveaux :

### 4.2.1 *La conception de l'architecture de l'innovation : le premier client à l'initiative du projet.*

Comme nous l'avons vu, le rôle de PressPro a concerné tout d'abord la constitution d'un partenariat entre éditeurs de logiciel : la nouvelle responsable du service documentation que nous qualifions d'« architecte innovation », a imaginé une solution agrégeant les logiciels de Temis et de Mondeca, et a sollicité ces éditeurs pour qu'ils associent leurs logiciels d'extraction terminologique et de gestion de base de connaissance. Ces derniers n'en ressentaient a priori pas la nécessité et se sont montrés réticents à son égard. Ensuite, c'est encore l'architecte innovation qui a suggéré d'associer Xyleme, spécialiste de l'archivage de documents au format XML. C'est elle, encore, qui a choisi de ne pas impliquer de société en charge de l'intégration de la solution, pensant que PressPro pourrait tenir ce rôle. Ainsi, l'acteur à l'initiative du projet a largement contribué à définir l'innovation. Son apport est venu également de la connaissance du métier des utilisateurs, de leur façon de travailler, et de leurs besoins non satisfaits en matière de traitement et d'archivage d'information.

Dans le projet Exinis, de la même manière, c'est un représentant de la DSI d'Exinis, appelé « architecte innovation », qui est à l'origine du projet et qui a sollicité Mondeca à son propos. Il a pu bénéficier de l'architecture de l'innovation conçue pour PressPro, mais n'en a retenu que les briques d'extraction d'information et de gestion de base de connaissance, pas celle, proposée par Xyleme, d'archivage de documents au format XML.

Le rôle d'un acteur, que nous appelons « architecte innovation » (ce qui est aussi la terminologie employée par les acteurs du projet) chez Exinis, et que l'on retrouve également chez PressPro, que nous pourrions aussi proposer de nommer « architecte pionnier » en référence à la traduction du terme « lead user » en « utilisateur pionnier », apparaît donc fondamental pour lancer le projet. Dans le contexte d'hyperspécialisation des start-up technologiques de l'industrie du logiciel, la constitution de solutions au travers de la combinaison de logiciels nécessitant le développement de composant logiciel d'interfaçage, peut difficilement être le fait d'un des éditeurs de logiciel impliqué. En effet, une telle démarche se heurte à cinq obstacles majeurs :

- Premièrement, la nécessaire focalisation sur sa technologie propre de la part de chaque éditeur de logiciel innovant, lui interdit de distraire des ressources significatives pour travailler à un tel couplage ;
- Deuxièmement, l'incertitude liée au marché d'une telle solution couplée, incite à la prudence quant à un tel investissement. Ce deuxième point est d'ailleurs souligné par von Hippel (1986, 1988) comme un des facteurs majeurs à l'origine de l'innovation par les utilisateurs. Il est ici pleinement vérifié, car l'idée originelle d'un couplage entre les logiciels de Temis et de Mondeca, était venue de la société Novartis, qui n'avait finalement pas donné suite. Par la suite, Temis et Mondeca n'ont pas travaillé sur le sujet, jusqu'à ce que PressPro relance un projet autour de ce thème. L'impulsion et le budget fournis par le client, sont donc bien ici les déclencheurs indispensables du processus d'innovation.
- Troisièmement, la sous-estimation de la spécificité technologique des autres éditeurs de logiciel : le projet PressPro montre des éditeurs qui ont, chacun, sous-estimé la spécificité de la technologie des autres éditeurs de logiciel. Mondeca a estimé que le moteur d'extraction était un composant logiciel, reposant sur une technologie certes nouvelle, mais suffisamment facile à maîtriser pour que la couplage avec sa base de connaissance soit aisé, et qu'il n'était nul besoin d'une implication forte de Temis dans le projet PressPro ; de même, Temis a estimé de son côté être capable de maîtriser la technologie des systèmes de représentation des connaissances, en tant qu'elle était moins radicalement nouvelle que celle de l'extraction terminologique, et qu'il lui serait donc facile, à terme, de la coupler avec son moteur d'extraction sans l'apport de Mondeca. Tandis que Temis et Mondeca estimaient pouvoir se passer de la technologie de Xyleme, dans le domaine de l'archivage de documents au format XML. Cela explique les flottements observés au début du projet PressPro, qui devait au départ être porté, pour l'aspect « couplage des logiciels cœur », par Mondeca seul, puis par Temis et Mondeca, et enfin, malgré les fortes réticences de ces deux firmes, par Temis, Mondeca et Xyleme sous l'injonction de l'architecte innovation de PressPro.
- Quatrièmement, le manque de compétences dans la gestion de projet impliquant un partenariat et une collaboration poussée avec d'autres éditeurs de logiciel innovants, qui se conjugue avec la méfiance ressentie envers des partenaires dont ils connaissent la fragilité, puisqu'elle est identique à la leur.
- Cinquièmement, le fait que l'information détenue par les utilisateurs soit « adhérente » (von Hippel, 1994). Cela rend difficile la compréhension a priori, pour le fournisseur, des

problèmes à résoudre et du bénéfice que les utilisateurs pourront retirer de leur résolution.

Ainsi, sans l'impulsion décisive, ni les choix, ni le financement d'un client, que l'on peut qualifier de « pionnier », nous n'avons pas la certitude qu'une solution innovante du type étudié, aurait pu émerger.

#### 4.2.2 *La conception des composants d'interfaçage entre les core products constitutants*

##### *L'innovation :*

Cette tâche a été effectuée par le chef de projet couplage assisté de l'ontologue expert de Mondeca, en collaboration avec le chef de projet cartouche de connaissance et les informaticiens de Temis. Cela a abouti aux différents composants logiciels qui constituent ce que nous avons appelé la passerelle OntoPop. Les clients, PressPro et Exinis, se sont donc bornés à fournir le terrain d'expérimentation du couplage envisagé, et à tester in fine, la qualité de son fonctionnement. Le manque de compétence des acteurs du projet chez PressPro, tant au niveau de l'architecte innovation, que des documentalistes ou encore des informaticiens chargés des problèmes de gestion documentaire leur a interdit de jouer un rôle actif dans la conception des ces composants d'interfaçage. Ce manque de compétence lié aux technologies d'extraction de l'information, et de gestion de base de connaissance est flagrant chez PressPro, aucun des acteurs engagés dans le projet n'ayant de formation ou d'expérience dans ces domaines. Il leur était alors impossible de jouer un rôle dans la conception d'un système reposant sur la combinaison de ces technologies, et donc au final sur la compréhension précise de chacune d'entre elles afin de pouvoir enrichir une base de connaissance ITM grâce aux extractions réalisées par le moteur IDE.

Le rôle joué par Exinis, quant à lui, dans la réalisation de cette tâche, a été beaucoup moins marginal. Le couplage ayant déjà été réalisé pour le premier projet, un premier prototype a pu être développé, et testé par Exinis. Exinis, par l'intermédiaire de son architecte innovation, a ainsi pu créer une première base de connaissance, assez basique, pour tester la capacité de l'ensemble de la solution à fonctionner, et donc du couplage IDE/ITM. Cela a permis de vérifier que ce couplage fonctionnait effectivement, et a mis en évidence des problèmes liés à la capacité de la base ITM, à fonctionner en tant que base relationnelle : notamment le temps d'accès à la base était trop long, à partir du moment où le volume de données à traiter dépassait une certaine limite. Les

compétences de l'architecte innovation et des deux utilisateurs experts, tant au niveau de leur métier (rédacteur juridique), que de l'analyse de données non structurées leur ont permis d'aider à concevoir cette plate-forme de test dès le début du projet, et de communiquer à Mondeca, principalement, des remarques précises, qui ont ensuite permis d'accélérer le fonctionnement de l'ensemble du système.

Ensuite, l'architecte innovation d'Exinis, s'appuyant sur les deux utilisateurs experts, a fait remonter le besoin de pouvoir consolider les lexiques du moteur d'extraction, par ce qui avait été extrait par ce même moteur, puis intégré à la base de connaissance. Il a persuadé Temis et Mondeca de concevoir un logiciel permettant de réaliser cette opération. D'ailleurs, ce logiciel a ensuite été intégré à la passerelle OntoPop dans le cadre du projet PressPro. Autre apport du client, l'utilisateur expert d'Exinis travaillant avec la chef de projet couplage, a formulé une demande relative à la création d'options dans les RAC, telles que la priorité ou le niveau de confiance accordée à une information. Ces apports qui peuvent être considérés comme mineurs, sont réels et ont constitué une amélioration dont l'intérêt a été jugé suffisant par Mondeca et Temis, pour qu'elle soit intégrée à la passerelle OntoPop elle-même. Dans le premier cas, c'est le résultat de la réflexion de l'architecte innovation, dans le deuxième cas celui de l'utilisateur expert travaillant sur les RAC.

En conclusion, le couplage de logiciels innovants apparaît comme la tâche la moins accessible aux acteurs du projet chez le client. Cela résulte de la nécessité de posséder un niveau de maîtrise élevé des deux technologies portées par ces logiciels, pour concevoir leur couplage. Dans un deuxième temps par contre, une fois qu'une première version de ce couplage a été élaborée, des apports de la part de ces acteurs sont envisageables, qui se sont avérés significatifs dès le deuxième projet le mettant en œuvre. Mais leur prise en compte dans le développement des logiciels concernés est restée le fait des éditeurs de logiciel, et non du client.

#### *4.2.3 La personnalisation des logiciels core products et de la passerelle OntoPop :*

La personnalisation concerne la création d'une cartouche de connaissance adaptée à un domaine précis, ensuite la définition de l'ontologie de ce domaine, et enfin la définition de RAC. Il s'agit donc bien ici de paramétrer ou de développer un composant complémentaire à un des logiciels, IDE, ITM, ou OntoPop. Dans le premier cas, cela exige des compétences de linguiste (i.e. dans le champs de l'analyse sémantique), dans le deuxième des compétences d'ontologie (i.e. dans le

champs de la représentation des connaissances), et le troisième une compréhension du fonctionnement de la base de connaissance ITM, et des résultats que peut fournir IDE : cela signifie à la fois une connaissance de ces logiciels (surtout ITM), et un certain niveau de connaissance en représentation des connaissances et en analyse sémantique, moins poussé toutefois que celles possédées par l'ontologue et le linguiste. Ces différentes personnalisations impliquent le client et les utilisateurs finaux à des degrés divers, comme nous allons l'analyser ci-après.

### **La cartouche de connaissance :**

Dans le projet PressPro, les utilisateurs n'ont pu participer à sa réalisation qu'au début du projet. Ensuite, Temis n'a interagi avec PressPro qu'au travers de l'architecte innovation. Son intervention s'est globalement bornée à conforter le chef de projet Temis et le chef de projet cartouche de connaissance de Temis dans leur travail. Finalement, le résultat présenté aux utilisateurs finaux n'a pas du tout donné satisfaction, et la cartouche de connaissance a été rejetée.

Dans le projet Exinis les deux utilisateurs experts d'Exinis ont travaillé de manière suivie avec le linguiste de Temis, après une première phase, courte, pendant laquelle un groupe d'utilisateurs classiques a été sollicité. Il s'agissait d'échanges permettant un transfert d'information sur le travail des utilisateurs. La cartouche de connaissance « juridique » est donc née réellement de l'interaction entre le linguiste de Temis et les deux utilisateurs experts. Le rôle des utilisateurs experts a été capital quant à la réussite de cette tâche, et il s'est inscrit dans le cadre d'une interaction suivie et approfondie entre client et fournisseur. Pour rendre telle interaction possible, le niveau de compétences et de motivation de ces utilisateurs experts apparaît fondamental. L'implication de ces utilisateurs, jugée indispensable à la fois l'ingénieur expert de Temis et par l'architecte innovation d'Exinis, montre le caractère « adhérent », *sticky*, de l'information détenue par les utilisateurs concernant leurs usages, et dont la connaissance est nécessaire au développement de l'innovation, ce qui correspond au point de vue de von Hippel sur la question (von Hippel, 1994).

### **L'ontologie :**

Dans le projet PressPro, l'ontologie a été traitée par le chef de projet ontologie de Mondeca, au travers d'ateliers utilisateurs. Après l'arrêt de ces ateliers, ils se sont poursuivis, d'une manière

beaucoup plus limitée, grâce à la participation de deux utilisateurs non experts dédiés à cette tâche, et travaillant dans les locaux de Temis. Il s'agissait pour le chef de projet ontologie de recueillir l'information nécessaire à l'organisation de la connaissance relative au secteur étudié, en tenant compte notamment des futures requêtes des utilisateurs.

Dans le projet Exinis, l'ontologie a été conçue par le chef de projet ontologie de Mondeca, avec le chef de projet intégration informatique, spécialiste des bases de données d'Exinis, et les deux utilisateurs experts, tous membres du comité de projet. Le résultat de leur travail conjoint était soumis à des groupes d'utilisateurs classiques, qui pouvaient y accéder au travers de l'application documentaire, qu'ils testaient aussi par la même occasion. Les utilisateurs classiques ont donc été sollicités, d'une manière très encadrée, mais leur apport a été jugé ici indispensable. Il est vrai que la définition de l'ontologie est très liée à l'utilisation de la base de connaissance par les utilisateurs finaux, et à la définition des interfaces utilisateur. Cette tâche nécessite donc l'apport des utilisateurs classiques, beaucoup plus que dans le cas de la cartouche de connaissance. Là encore, cela illustre la nécessité du transfert d'informations « adhérentes » que seuls possèdent les utilisateurs, afin que les acteurs en charge du développement de l'innovation puissent opérer. La différence, avec la cartouche de connaissance, vient de la consultation régulière des utilisateurs classiques au travers d'ateliers. Cela illustre, à notre sens, le fait que la définition de l'ontologie est directement liée aux usages des rédacteurs d'Exinis. En conséquence cette définition requiert la prise en compte d'une grande variété d'usages et d'opinions, qu'il s'agit d'observer ou de recueillir auprès d'utilisateurs classiques. C'est donc une situation un peu différente de celle observée dans le cas de la cartouche de connaissance, où l'apport des utilisateurs classiques apparaît moins utile.

Le projet Exinis propose une approche en deux temps, une collaboration suivie entre le chef de projet ontologie et les utilisateurs experts, qui donne lieu à des validations ponctuelles par des utilisateurs classiques. Cette tâche résulte de l'interaction client/fournisseur, qui est elle-même facilitée par le niveau de compétences et la motivation des utilisateurs appelés à collaborer avec le chef de projet ontologie.

### **Personnalisation des RAC :**

Dans le projet PressPro, la chef de projet couplage a d'abord travaillé au développement de la passerelle OntoPop, avant de travailler à la définition de RAC personnalisées. Ce travail fondateur s'est achevé fin 2005, et cette personnalisation a ensuite été occultée par le problème de

la cartouche de connaissance, qui ne donnait pas satisfaction aux utilisateurs de PressPro. Le lien entre IDE et ITM au travers des RAC, n'a donc pas pu être testé auprès des utilisateurs de PressPro.

Dans le projet Exinis, la chef de projet couplage a travaillé avec un des utilisateurs experts, mais pas directement avec les autres utilisateurs. Cet utilisateur expert a travaillé avec le comité de projet, puis avec les utilisateurs classiques au travers des mêmes types d'atelier que pour l'ontologie. En conclusion sur ce point, le rôle des utilisateurs s'est limité à la collaboration d'un des utilisateurs experts avec la responsable de couplage, d'une manière régulière et suivie. L'interaction avec certains utilisateurs experts apparaît là aussi comme indispensable à cette personnalisation, alors que l'apport des utilisateurs classiques est comparable à celui observé pour la définition de l'ontologie. On note enfin que l'architecte innovation d'Exinis, dans la perspective de « prendre le contrôle de la solution », a cherché à ce que les informaticiens d'Exinis apprennent à modifier des RAC en utilisant l'éditeur de RAC fourni par Mondeca. Mais cette volonté n'a pu aboutir, l'expérience montrant que la modification de RAC n'est pas aisée, même après une formation par la chef de projet couplage, si la personne en charge de cette tâche n'a pas elle-même participé à leur conception. Cet exemple montre la volonté d'appropriation de l'innovation par l'architecte innovation d'Exinis, et les difficultés aux quelles cette volonté se heurte.

#### *4.2.4 La conception de l'application documentaire :*

Les clients ont joué un rôle clef dans la réalisation (conception et développement) de l'application documentaire, qui définit les interfaces utilisateur.

- **Dans le projet PressPro** : Le développement de l'application documentaire, a d'abord été attribué aux informaticiens du service documentation de PressPro. Mais le manque de compétences relatives aux technologies nouvelles, en général, et aux logiciels employés en particulier, a constitué une difficulté infranchissable par les informaticiens de PressPro. Cette difficulté fut d'autant plus infranchissable qu'elle était sous-estimée par les éditeurs du consortium eux-mêmes. Ainsi, il est apparu que la nécessaire proximité avec les utilisateurs, l'indispensable compréhension de leur façon de travailler, ne suffit pas : il faut être capable de leur expliquer ce que peut leur apporter le système, et dans une certaine mesure la façon avec laquelle ce système allait fonctionner, pour ensuite pouvoir recueillir les informations

indispensables au développement à réaliser de leur part. Il faut aussi être capable de comprendre le projet dans sa globalité et d'interagir avec Temis, Mondeca et les autres éditeurs de logiciels impliqués dans le projet de création de base de connaissance.

Face à ce constat d'échec, c'est encore PressPro qui a contraint Temis et Xyleme à assumer cette tâche malgré leurs réticences persistantes. Mais chacun dû convenir qu'il n'y avait pas de SSII susceptibles de la reprendre. Ce développement a été principalement conçu et développé par le chef de projet de Xyleme, secondé par les informaticiens de PressPro dans un premier temps, puis par un informaticien spécialement recruté par Xyleme, pour la partie développement proprement dite. La conception de l'application documentaire s'est appuyée sur des ateliers utilisateurs, où les documentalistes décrivaient leur façon de travailler, expliquaient ce qu'elles désiraient, et finalement réagissaient aux propositions qui leur étaient faites.

- **Dans le projet Exinis** : Face aux certitudes acquises sur ce sujet par Temis et Mondeca, l'architecte innovation d'Exinis, a convenu, dès le départ, de l'intérêt de découpler les tâches de personnalisation des composants du système, de celle de développement de l'application documentaire. Il s'est alors agi pour lui, de choisir une société de services pour l'exécuter, Temis et Mondeca ayant insisté sur le fait qu'ils ne prendraient pas en charge cette tâche. L'architecte innovation a retenu une société de services de petite taille spécialisée dans les systèmes de gestion documentaire et ayant déjà travaillé pour Exinis. L'obstacle de la connaissance du « métier » d'Exinis, selon l'expression employée par les acteurs du projet, était alors levé. De même, la capacité à contrôler l'activité de cette société de services était facilitée par les rapports antérieurs entre celle-ci et Exinis, tout comme elle l'était par l'importance du client Exinis, d'un point de vue général, pour la société 4D Concept : de ce fait, l'architecte innovation voyait ses doutes quant à la motivation de la société de services, et quant à la difficulté de contrôler son action, levés. Toutefois, cette organisation suscitait des doutes auprès des dirigeants d'Exinis France, et ils exigèrent qu'une SSII de grande envergure soit impliquée dans le projet. Unilog se vit ainsi confier un rôle de superviseur global du projet, et très concrètement la tâche d'organiser les ateliers au cours desquels des utilisateurs classiques étaient sollicités. Mais tout comme les informaticiens de PressPro, n'ont pu que rester étranger au projet précédent, le chef de projet d'Unilog n'a pu apporter de contribution significative, n'a pu trouver de place dans celui-là, et Unilog a rapidement quitté le projet Exinis.

**En conclusion**, la réalisation de l'application documentaire touche directement les utilisateurs pour lesquels les interfaces doivent être définies. Elle demande une compréhension précise de leur façon de travailler, et sert d'interface entre les utilisateurs et la base de connaissance, qui peut ensuite solliciter d'autres composants logiciels, comme un moteur d'extraction terminologique, ou un système de gestion de contenu (projet PressPro, avec XML Server). Elle peut aussi appeler des web services, permettant, par exemple, la consolidation d'informations particulières (projet Exinis). La nécessité de la prise en compte d'un maximum d'usages est encore plus marquée dans le cas de l'application documentaire que dans celui de l'ontologie ou des RAC, car cette application constitue l'interface utilisateur du système. Ainsi, les tests utilisateurs portant sur l'application documentaire, peuvent permettre de faire remonter des problèmes de couplage, des problèmes de définition d'ontologie, des problèmes de requêtes, entre autres. C'est pourquoi l'implication d'utilisateurs expert, mais aussi d'utilisateurs classiques est aussi importante ici. L'implication de l'architecte innovation, doté d'une vision globale et susceptible de faire remonter chaque problème aux acteurs concernés, est elle-aussi capitale. Le rôle du client est donc ici multiple. D'abord, les utilisateurs classiques et experts doivent, en phase de développement, décrire et expliquer leur manière de travailler ; ensuite ces utilisateurs doivent tester la solution, et faire émerger des problèmes qui peuvent provenir de n'importe quel composant. C'est pourquoi l'architecte innovation a un rôle à jouer, en tant qu'acteur possédant une vision globale de cette solution, et susceptible d'intervenir auprès des autres acteurs afin de s'assurer de la prise en compte de chacun des problèmes ainsi mis en évidence.

Plus globalement, nous pouvons faire l'hypothèse que la tâche de réalisation de l'application documentaire, ne peut généralement pas être attribuée au client, car elle requiert des compétences en termes de développement d'interfaces utilisateur que celui-ci ne possède pas systématiquement.

#### *4.2.5 La coordination globale du projet :*

Le client a joué un rôle dans la coordination globale des projets PressPro et Exinis. Le client PressPro a tenté de jouer ce rôle de différentes manières :

- en refusant qu'une société de services assume ce rôle,

- en choisissant de petits éditeurs de logiciel indépendants, pour lesquels le projet constituait un enjeu significatif (financièrement, technologiquement, et commercialement), et que PressPro estimait pouvoir contrôler,
- en choisissant de faire développer l'application documentaire en interne,
- en structurant des groupes d'utilisateurs.

Force est de constater que ce client n'a pas réussi à jouer pleinement ce rôle. C'est certainement un des éléments qui ont conduit à l'échec final, comme nous l'avons déjà évoqué. Toutefois, la question se pose de savoir qui aurait été en mesure de l'assumer.

Dans le projet Exinis, la direction et donc la coordination globale, du projet est en grande partie revenue à l'architecte innovation. A la différence de l'architecte innovation de PressPro, il était pratiquement dédié au projet. Il a joué un rôle fondamental à ce niveau-là, mais il ne l'a pas joué seul :

- Il a supervisé les différentes tâches :
  - La conception de la cartouche de connaissance,
  - La conception de l'ontologie,
  - le couplage personnalisé entre IDE et ITM,
  - Et surtout, le développement de l'application documentaire par 4D Concept.
- Il a été une force de proposition avec les deux utilisateurs experts, grâce à sa vision globale du projet et de ses développements futurs.
- Il a défendu le projet auprès des directions d'Exinis France, Europe et International. A cet effet, il a inclus dans le planning de développement de la solution, des objectifs intermédiaires susceptibles de donner des exemples concrets aux managers d'Exinis, de la valeur ajoutée que pourrait apporter la solution.

Les personnes à l'origine du projet chez PressPro et chez Exinis, les « architectes innovation » de ces deux projets, ont donc joué un rôle capital tout au long des deux projets étudiés, notamment quant à la coordination globale du projet, mais avec des résultats différents. A ce stade, il nous apparaît que l'étude de leurs profils et parcours respectifs est susceptible d'apporter un éclairage quant à leur contribution à ces projets :

- - L'architecte innovation de PressPro a reçu une formation de documentaliste. Elle possédait une expérience de documentaliste dans différents groupes de presse de 1973 à 1990, puis de chef de documentation dans un journal économique de 1990 à 2002, avant de devenir chef de documentation de PressPro à partir de 2002.
- - L'architecte innovation d'Exinis quant à lui, est ingénieur. Il possède une expérience de chercheur en informatique, qui l'a amené à collaborer avec des chercheurs en traitement du langage ; il est ensuite entré en 2000, au service informatique d'Exinis France, avec pour mission principale de transférer sur Internet le contenu qu'Exinis France mettait à la disposition de ses clients.

L'architecte innovation d'Exinis, possédait donc une expérience de la recherche en informatique et des liens avec des chercheurs en représentation des connaissances, à la différence de l'architecte innovation de PressPro. Cette dernière possédait par contre une longue expérience de documentaliste et de responsable de documentation dans des groupes de presse. Ce type d'expérience, d'« utilisatrice » ne lui a pas permis de coordonner efficacement le projet, à la différence de l'architecte innovation d'Exinis, qui lui n'était pas un « utilisateur », mais possédait une vision globale des technologies employées et de leur apport potentiel, grâce au caractère transversal de ses compétences. Il avait aussi, toujours à la différence de l'architecte innovation de PressPro, déjà collaboré avec des chercheurs dans les champs d'expertise de Temis et de Mondeca, ce qui a, de toute évidence, favorisé sa capacité à diriger le projet.

### 4.3. Synthèse des rôles joués au cours des projets PressPro et Exinis :

En synthèse, nous reprenons dans le tableau 6 ci-dessous les différentes tâches évoquées au 3.2.1.

	<b>Projet PressPro</b>	<b>Projet Exinis</b>
<b>Conception de l'architecture de l'innovation</b>	<b>chef du service documentation de PressPro (architecte innovation).</b>	<b>Informaticien de la DSI d'Exinis (architecte innovation).</b>
<b>Intégration informatique de la solution logicielle</b>	<b>Responsable informatique de la documentation de PressPro (chef de projet intégration informatique), puis</b> Chef de projet Xyleme.	<b>Informaticien de la DSI spécialiste des bases de données d'Exinis (chef de projet intégration informatique).</b>
<b>Conception et réalisation des modules de couplage</b>	Chef de projet couplage de Mondeca, assistée du service informatique de Mondeca, en collaboration avec le linguiste de Temis et l'ontologue de Mondeca	Modules déjà conçus pour le projet PressPro, mais ajout d'une partie « options » et « consolidation » par la chef de projet couplage de Mondeca, suite aux demandes d'un <b>utilisateur expert d'Exinis</b> .
<b>Conception et réalisation des composants de personnalisation</b>	<p>- <b>Cartouche de connaissance</b> : Linguiste (chef de projet cartouche de connaissance) de Temis, en collaboration avec <b>un groupe d'utilisateurs classiques</b> au travers d'ateliers suspendus après quelques mois ; ensuite travail autonome du linguiste de Temis et échanges ponctuels avec <b>l'architecte innovation</b>.</p> <p>- <b>Ontologie</b> : Ontologue de Mondeca en collaboration avec <b>un groupe d'utilisateurs classiques au travers d'ateliers</b>, remplacé ensuite par <b>deux utilisateurs classiques dédiés</b>.</p> <p>- <b>RAC</b> : Chef de projet couplage de Mondeca, avec le linguiste de Temis et l'ontologue de Mondeca.</p>	<p>- <b>Cartouche de connaissance</b> : Linguiste et responsable du projet de Mondeca en contact avec <b>un groupe d'utilisateurs</b>, puis uniquement avec <b>les deux utilisateurs experts</b> d'Exinis ;</p> <p>- <b>Ontologie</b>: Ontologue (chef de projet ontologie) et le chef de projet de Mondeca avec <b>le chef de projet intégration informatique et les utilisateurs experts</b> ;</p> <p>- <b>RAC</b> : Chef de projet couplage de Mondeca avec un <b>utilisateur expert</b>.</p> <p>[Réunions « ateliers utilisateurs» entre le comité de projet Exinis et des utilisateurs classiques (15 rédacteurs) : pour chacune de ces tâches (sauf cartouche de connaissance).]</p>
<b>Conception et réalisation de l'application documentaire</b>	<ol style="list-style-type: none"> <li><b>Informaticiens de PressPro, puis</b></li> <li>Chef de projet de Xyleme et un informaticien de Xyleme, en collaboration avec le chef de projet de Mondeca, <b>la chef de documentation et les deux utilisateurs dédiés</b>.</li> </ol>	Chef de projet de 4D Concept avec les <b>deux utilisateurs experts</b> , ces derniers interagissant avec des <b>utilisateurs classiques</b> , et avec le support de <b>l'architecte innovation</b> .

Tableau 6 : Rôles joués par les acteurs situés chez le client (en gras) et les autres acteurs dans le processus d'innovation.

Nous ajoutons à cette synthèse, les rôles joués au niveau de la structuration de ces projets (tableau 7) :

	<b>Projet PressPro</b>	<b>Projet Exinis</b>
<b>Coordinateur global</b>	Architecte innovation assistée du chef de projet intégration informatique de PressPr <i>[pas de coordination globale en pratique].</i>	Architecte innovation assisté du chef de projet intégration informatique d'Exinis.
<b>Coordinateur du couplage</b>	Chef de projet couplage de Mondeca.	Chef de projet couplage de Mondeca.
<b>Coordinateurs application documentaire/ontologie</b>	Chef de projet Xyleme et chef de projet ontologie de Mondeca.	Chef de projet de 4D Concept assisté du chef de projet ontologie de Mondeca.

**Tableau 7 : Rôles joués par les acteurs situés chez le client et par les autres acteurs dans la structuration des projets.**

#### 4.4 Enrichissement de la notion de *lead user* :

##### 4.4.1 Des caractéristiques de lead users:

Nous pouvons considérer que les clients PressPro et Exinis présentaient des caractéristiques de *lead users*, car d'une part ils étaient à la pointe d'une tendance nouvelle du marché des logiciels de traitement de document (le *text mining*), et car d'autre part, ils espéraient retirer des bénéfices substantiels de l'utilisation des nouveaux produits qui apparaissaient dans ce domaine. Tout d'abord, ces clients recherchaient des solutions à leurs problèmes spécifiques. L'innovation en satisfaisant leur besoin devait leur apporter des gains : gain de productivité et enrichissement des tâches des utilisateurs pour le projet PressPro, offre enrichie pour le projet Exinis, génératrice de valeur ajoutée pour ses clients. De plus, il nous semble que les besoins de ces clients étaient à la

fois spécifiques et à l'avant-garde d'une tendance d'un marché, l'analyse automatisée de texte « non structuré », jugée prometteuse par un certains nombres d'experts. En effet, le volume croissant d'information accessible aux organisations et individus, pouvait laisser espérer une explosion de ce marché, similaire à celle du marché du *data mining* (analyse et le traitement d'informations formatées), des années auparavant.

Mais qui est ou qui sont le(s) *lead user(s)* dans les deux projets analysés ? Pour von Hippel (1988, 2005), il peut s'agir d'un individu, du département d'une entreprise, ou d'une entreprise toute entière. Dans notre cas, il vient immédiatement à l'esprit que les architectes innovation de PressPro et d'Exinis, qui sont à l'origine des projets, pourraient être considérés comme des *lead users*: à la pointe d'une tendance nouvelle, le *text mining*, ils estiment que l'intégration de cette technologie, sera profitable à leur entreprise. De même, chez Exinis, les deux utilisateurs experts, dotés d'une formation en gestion des connaissances, présentent des caractéristiques de *lead users*, sensibilisés, plus que formés au *text mining*, ils comprennent comment cette technologie peut augmenter la valeur ajoutée du travail qu'ils réalisent pour les clients finaux. Ainsi, il apparaît qu'il existe chez les clients PressPro et Exinis, différents acteurs ou groupes d'acteurs présentant les caractéristiques de *lead users*, « individuels », c'est pourquoi nous considérons que les clients PressPro et Exinis constituent des *lead users* « collectifs ». Nous reviendrons toutefois sur cette notion, pour proposer des notions complémentaires susceptibles de s'appliquer à différents acteurs ou groupes d'acteurs jouant un rôle dans le processus d'innovation.

#### 4.4.2 Nature des rôles joués par les acteurs clefs des projets chez PressPro et Exinis :

Parmi les rôles que nous avons listés, certains sont assumés par un acteur ou groupe d'acteurs situé chez le client, d'autres par les fournisseurs, et d'autres encore ne peuvent se réaliser qu'au travers de l'interaction entre le client et un ou plusieurs de ses fournisseurs :

- La définition de l'architecture nouvelle revient clairement au client,
- Le pilotage global, qui inclut la coordination globale, du projet revient au client.
- La réalisation des modules couplés revient aux fournisseurs,
- La réalisation des composants personnalisés et des paramétrages suppose une interaction entre les fournisseurs concernés et des utilisateurs experts,

- La réalisation de l'interface utilisateur, l'application documentaire, résulte de l'interaction entre les mêmes acteurs qu'au point précédent, auxquels il faut ajouter des utilisateurs classiques,

Il apparaît donc que le rôle joué dans un processus d'innovation radicale modulaire multi-acteurs, par un client possédant des caractéristiques de *lead users*, peut rencontrer certaines limites.

#### 4.4.3 Enrichissement de la notion de lead user :

##### 4.4.3.1 Introduction :

Nous sommes au terme de ce travail en mesure de proposer un enrichissement de la notion de *lead user*, définie par von Hippel (1986). Nous soulignerons tout d'abord, les points qui nous semblent particulièrement pertinents dans l'approche préconisée par la littérature sur l'innovation par les utilisateurs, puis nous insisterons sur la multiplicité des rôles joués par le client dans le processus d'innovation, et la multiplicité des acteurs qui, chez le client, joue un rôle. Nous proposerons de nouvelles terminologies permettant de qualifier ces différents rôles. Enfin, nous soulignerons le fait que, certains lieux de capitalisation et certaines compétences ne se trouvant pas chez le client, certaines tâches ne peuvent être assumées par le client seul.

##### 4.4.3.2 L'indispensable contribution du client au processus d'innovation :

Tout d'abord nous constatons que le rôle du client, ou plutôt de différents acteurs situés chez le client, est fondamental dans le processus de l'innovation quasi-modulaire multi-acteurs étudiée. Ce rôle est prégnant dès l'origine, l'architecture de l'innovation étant imaginée par l'architecte innovation de PressPro. Elle parvient à imposer sa vision de cette architecture nouvelle, à des éditeurs de logiciel réticents à l'idée d'y contribuer. Leur comportement s'explique tout d'abord par le fait qu'ils ne possèdent pas la vision globale de la solution recherchée par le client, ni la maîtrise des technologies employées par leurs futurs partenaires, ni l'expérience de projets nécessitant des coopérations approfondies entre éditeurs de logiciel. De même, on constate que le premier client est prêt à financer la conception d'une innovation alors que d'eux-mêmes Temis et Mondeca y avaient renoncé, après que Novartis avait une première fois, envisagé le couplage entre IDE et ITM. Pour eux, le marché étant très incertain, l'investissement requis par ce couplage était trop risqué, alors que pour PressPro, la connaissance des usages futurs, et des gains de productivité réalisables grâce à l'innovation, constituait une motivation suffisante. Le risque

d'innover a donc été pris par le client plutôt que par les fournisseurs. C'est d'autant plus vrai que les acteurs sont multiples et focalisés sur leur propre technologie : tout investissement hors de leur sphère technologique propre, i.e. tout investissement qui n'est pas consacré à leur technologie cœur, aboutit à une perte de temps et de ressource par rapport à leurs concurrents dans leur domaine technologique. Dans le secteur du logiciel, nous avons ainsi vu (cf. chapitre 1 sur l'industrie du logiciel) à quel point il est important de se positionner parmi les premiers acteurs d'un nouveau domaine applicatif pour avoir des chances d'y réussir commercialement.

Nous pouvons en conclure que des éléments objectifs ont contribué à placer le client, en tout cas certains d'entre eux, dans une position plus favorable que les fournisseurs concernés, pour imaginer l'architecture et lancer, en la finançant, la conception d'une innovation reposant sur une combinaison nouvelle de logiciels eux-mêmes innovants. Cela va dans le sens d'une confirmation des théories de von Hippel relatives à l'innovation par les utilisateurs (1986, 1988, 2005), notamment dans l'industrie du logiciel avec les exemples des logiciels PC-CAD et Apache.

#### 4.4.3.3 Une contribution multiple du client, par des acteurs divers :

##### ***La conception de l'architecture de l'innovation :***

La conception de cette architecture revient à la chef de documentation de PressPro, que nous avons appelé architecte innovation, sa mise en œuvre dans un contexte très exigeant et son enrichissement pour une grande part à l'architecte innovation d'Exinis. Nous venons d'exposer plus haut, les raisons du rôle du client concernant cette tâche. Nous proposons de nommer l'acteur qui assume le rôle chez le client, client considéré dans sa globalité comme un *lead user*, « **lead user architecte** ». Ce « *lead user architecte* » est en position d'assumer la fonction de directeur de projet (Midler, 1993 ; Giard et Midler, 1993) ou « heavy weight project manager » (Clark et Wheelwright, 1992), pour différentes raisons qui ont été mises en évidence lors de notre travail de recherche : tout d'abord, il possède une légitimité interne et externe, parce qu'il est celui qui a imaginé l'architecture de l'innovation, bien sûr, mais aussi parce qu'il est celui qui a du et qui a su, convaincre le management de son entreprise de l'intérêt de cette innovation. Cette légitimité et cette capacité à imaginer une architecture nouvelle le conduisent logiquement à assumer le rôle de directeur de projet, qui oriente, supervise et assume la coordination globale du projet, mais qui doit aussi rendre des comptes à son management. Cela nécessite à la fois des compétences transversales, des moyens et du temps : on comprend alors que l'architecte innovation d'Exinis ait été dans une situation nettement plus favorable que l'architecte innovation de PressPro pour

assumer ce rôle. En tant qu'architecte de l'innovation et directeur de projet, il peut être considéré comme le maître d'ouvrage du projet,

### ***L'apport des utilisateurs classiques :***

On trouve au démarrage des deux projets étudiés des ateliers utilisateurs, au cours desquels les acteurs concernés chez Temis et Mondeca, ont cherché à comprendre comment les utilisateurs « classiques », qui ne possédaient ni compétence ni motivation particulières relatives à l'innovation, travaillaient, et comment ils pourraient s'approprier un outil logiciel innovant. Il s'agissait ici de faire contribuer un assez grand nombre d'utilisateurs classiques, principalement dans le but d'observer le maximum d'usages, et de les faire réagir à des propositions concrètes relatives à l'application en développement. A partir de l'analyse de ces usages et de ces réactions, des composants logiciels personnalisés ont été développés, et les paramétrages de logiciel réalisés. Cet apport indispensable des utilisateurs classiques trouve son origine dans le caractère « adhérent » des informations dont il dispose à propos de leur propre activité, caractère « adhérent » dont les projets que nous avons étudiés montrent à quel point il est marqué, confirmant en cela une des hypothèses fondatrices de l'approche *user innovation* (von Hippel, 1994). Nous proposons de nommer les utilisateurs « classiques » d'un client possédant les caractéristiques de *lead users*, « ***lead users utilisateurs classiques*** ».

### ***Le développement des composants personnalisés avec des lead users « individuels » :***

Ce développement est passé, au cours du projet Exinis, par la collaboration entre des experts de Temis ou de Mondeca, et des utilisateurs, que nous avons appelés « utilisateurs experts ». Cette collaboration a été régulière et suivie, les utilisateurs qui y ont participé étaient dédiés à cette tâche. Dans le projet Exinis, il s'agissait d'utilisateurs disposant d'un niveau de compétence facilitant leur collaboration avec Temis, Mondeca ou 4D Concept, concernant la conception des différents composants (modules ou quasi-modules) personnalisés. De plus, il s'agissait d'utilisateurs présentant des caractéristiques de *lead users*, car ils étaient d'une certaine manière « à la pointe » de la tendance du *text mining*, de par leur intérêt et leur formation dans un champ connexe, et parce qu'ils comprenaient ce que cette technologie pouvait leur apporter dans le cadre de leur activité. Le projet PressPro nous a montré, quant à lui, une collaboration qui a impliqué deux utilisateurs (suite à l'arrêt des ateliers utilisateurs), qui au départ ne possédaient ni expérience, ni compétences particulières dans les domaines qui nous intéressent. La collaboration suivie avec le chef de produit ontologie de Mondeca et le chef de projet de Xyleme, a constitué

pour eux un apprentissage, qui progressivement leur a permis de mieux interagir avec les fournisseurs, dans une certaine limite.

En conclusion, nous retenons le rôle capital de ces utilisateurs, qui possèdent des compétences et une motivation qui les distingue des utilisateurs classiques, dans le processus d'innovation. Ces compétences facilitent la transmission des informations « adhérentes » aux acteurs en charge du développement de l'innovation, et plus globalement permettent l'instauration d'un dialogue fécond entre ces utilisateurs et ces acteurs. Ces utilisateurs compétents qui collaborent avec les acteurs en charge du développement, peuvent aussi jouer le rôle d'interface entre ces acteurs et les utilisateurs classiques. Nous avons vu lors de notre étude, que ce rôle était à la fois crucial et difficile à assumer. Nous proposons de nommer les utilisateurs qui possèdent des compétences relatives aux technologies employées par l'innovation et qui possèdent les caractéristiques de *lead user*, « **lead users experts** ».

#### ***La gestion opérationnelle du projet :***

Dans le projet Exinis, le chef de projet intégration informatique a géré de manière opérationnelle le projet, il est ainsi amené à attribuer, suivre et contrôler les différentes tâches. Il joue une grande partie du rôle de maître d'œuvre, de « chef de projet de réalisation » (Charue-Duboc, 1997), qui est responsable de la réalisation de l'ouvrage et qui doit mettre en œuvre les moyens techniques nécessaires définis au préalable par le maître d'ouvrage. Son rôle est indispensable à la bonne marche quotidienne du projet, c'est un gestionnaire de projet qui a un statut beaucoup plus faible que celui du directeur de projet, qu'il assiste. Dans les projets étudiés, ce rôle a été tenu par les acteurs que nous avons nommés « chef de projet intégration informatique ». Cet acteur se situe logiquement chez le client, notamment parce que les différents fournisseurs constituent des organisations faiblement couplées, qui plus est focalisées sur leur technologie propre, et, en dehors du cas particulier du couplage moteur d'extraction/base de connaissance, c'est bien le client qui est le mieux placé pour interagir avec chacun d'entre eux, et s'assurer de la cohérence des différentes contributions. Ainsi, ce rôle indispensable de « chef de projet de réalisation », doit-il être assumé par le client *lead user*. Nous choisissons d'appeler cet acteur « **lead user gestionnaire de projet** ». Les compétences requises pour assumer un tel rôle semblent moins étendues que dans le cas du « *lead user* architecte », car il ne s'agit pas ici d'imaginer des solutions nouvelles, ni d'orienter le projet, mais de le comprendre et de s'assurer de son bon déroulement sous la supervision du directeur de projet. Cela signifie donc tout de même une capacité à interagir avec chacun des acteurs du projet, et donc à posséder aussi des compétences

transversales, ainsi bien sûr que des compétences concrètes en gestion opérationnelle de projet de systèmes d'information.

#### 4.4.4 Conclusion sur la contribution du client au processus d'innovation :

Différents acteurs chez le client jouent différents rôles, architecte innovation, gestionnaire de projet, utilisateur, expert concepteur, utilisateur classique dans le processus d'innovation mais, aucun acteur n'est en mesure de tenir tous ces rôles à la fois. Les différents rôles joués par les clients ne parviennent pas à couvrir l'ensemble des tâches du processus d'innovation: certaines tâches revenant uniquement aux fournisseurs, d'autres ne pouvant être réalisées qu'au travers d'une collaboration client/fournisseur.

L'innovation étudiée n'aurait pas pu exister sans l'action des deux premiers clients, pour imaginer l'architecture, pour expliquer les usages, pour orienter et coordonner les projets. Mais le rôle joué par le client, a été éclaté entre plusieurs acteurs, parmi lesquels certains possédaient les caractéristiques d'individus *lead users*. Ces individus *lead users* jouent effectivement un rôle capital dans le processus d'innovation, mais ils ne sont donc pas les seuls. Les « individus *lead users* » ont apporté une contribution unique au processus d'innovation. Il est possible de les caractériser par leur niveau de compétence et leur ouverture aux technologies nouvelles qui fondent l'innovation. L'absence d'utilisateurs possédant des caractéristiques de « *lead user expert* » a pénalisé le projet PressPro, tandis que la présence de deux rédacteurs juridiques experts possédant ces caractéristiques chez Exinis, a été exploitée au maximum par Temis, Mondeca et 4D Concept, qui ont reconnu que ce binôme a joué un rôle crucial dans la réussite du projet.

L'apport des utilisateurs classiques situés chez un client *lead user*, que nous appelons « *lead user* utilisateurs classiques », qui sont très largement majoritaires, au travers d'ateliers utilisateurs apparaît également indispensable. Il s'agit ici de tenter de saisir tous les usages possibles, mais aussi de faire la pédagogie de l'innovation pour éviter un rejet ultérieur. Il apparaît donc clairement que les rôles joués ici par les utilisateurs classiques ne peuvent pas être assumés par quelques utilisateurs dotés du profil de *lead user*, car il s'agit d'une part d'observer le maximum d'usages possibles, et le nombre d'utilisateurs impliqués importe, et d'autre part de convaincre le maximum d'utilisateurs de l'intérêt de l'innovation. Le premier point renvoie d'ailleurs en partie au rôle des communautés d'utilisateurs, que ce soit dans le domaine du logiciel propriétaire ou

dans celui du logiciel Open Source (Raymond, 1999 ; Franke et von Hippel, 2003 ; von Krogh et von Hippel, 2003).

Mais cette interaction avec les utilisateurs classiques est plus difficile à gérer par les éditeurs de logiciel concernés, comme le montre l'échec des ateliers utilisateurs de PressPro. De même, l'absence d'interaction de ce type pendant toute la deuxième moitié du projet Exinis s'explique par la crainte du comité de projet, principalement l'architecte innovation, que les éditeurs de logiciel ne sachent pas la gérer. C'est pourquoi le dispositif qu'il met en place, consiste avant tout à faire interagir ces utilisateurs avec les « *lead user experts* », qui de fait jouent un rôle d'interface, facilitant la compréhension et l'appropriation de l'innovation par la masse des utilisateurs.

#### **4.5 Des tâches qui ne peuvent être assumées par le client seul :**

##### **Des tâches qui nécessitent une interaction fournisseur client :**

Certaines tâches ne peuvent être assumées entièrement par des acteurs situés chez le client, parce qu'elles requièrent des compétences rares que seules certaines startups logicielles innovantes possèdent et développent. Mais certaines de ces tâches nécessitent aussi une interaction client/fournisseur permettant des échanges dialectiques relatifs aux usages, voire au système d'information du client. L'exécution de ces tâches passe comme nous l'avons vu par l'interaction entre le représentant d'un éditeur de logiciel et des *lead users*. A terme, un transfert de compétence est envisageable, aboutissant à la réalisation de telles tâches par le client seul, comme la réalisation de cartouche de connaissance ou la définition de nouvelles RAC. Dans le deuxième cas, l'outil logiciel existe (l'« éditeur de règles »), tandis que dans le premier, il reste à Temis à développer un outil logiciel convivial et documenté.

##### **Des tâches qui ne peuvent être assumées que par les fournisseurs :**

Certaines tâches n'ont été, et ne pouvaient être assumés que par Temis ou Mondeca. C'est le cas du couplage entre moteur d'extraction terminologique et système de représentation des connaissances, qui a donné lieu à la création de la passerelle logicielle OntoPop, d'une compétence nouvelle, celle de « chef de projet couplage », et à l'établissement d'une méthodologie. Toutes choses qui ont été définies par Mondeca, avec l'aide de Temis.

Nous voyons à cela principalement deux raisons : les compétences et la motivation. D'une part, comme nous l'avons déjà dit, il s'agissait de coupler deux technologies nouvelles non maîtrisées par le client. Il fallait donc posséder en interne des experts, linguiste ou ontologue, dont les compétences étaient rares, ce qui était le cas de Temis et de Mondeca. D'autre part, le premier projet avait dès son démarrage convaincu Temis et Mondeca de la viabilité et du potentiel commercial d'une plate-forme logicielle (IDE/OntoPop/ITM), sur laquelle peuvent venir se brancher des modules tels que XML Server de Xyleme, des Webservices de consolidation de l'information, des moteurs d'inférence ou des moteurs de recherche, comme ce sera le cas dans des projets ultérieurs. Cette motivation forte les a poussés à investir et à recruter des personnes dédiées à ce couplage. Les premiers clients ne possédaient pas de telles compétences, ni une telle motivation, c'est pourquoi ils ne pouvaient assumer les tâches ayant trait à ce couplage. Par la suite, la capitalisation relative à celui-ci ne s'est faite pas chez le client, contrairement à la vision exprimée par la littérature sur l'innovation par les utilisateurs. Ainsi, le client Exinis qui souhaitait depuis le début s'approprier la solution n'avait pas encore pu y parvenir au terme de notre recherche. L'architecte innovation voulait en effet pouvoir assurer la maintenance de la cartouche de connaissance juridique et modifier les RAC définies par Mondeca. Il voulait ensuite rapidement utiliser le *toolkit* de développement de la cartouche de connaissances de Temis. Malgré cette volonté et les moyens dont il disposait, aucune de ces trois tâches n'était encore assumée par Exinis mi-2008, alors que l'application était en production. On voit ici qu'un haut niveau de motivation ne suffit pas à compenser un niveau de compétences limité.

Dans ces conditions, l'émergence d'une communauté d'utilisateurs susceptible d'échanger à propos de l'innovation, n'est pas envisageable à court terme.

## **5. CONCLUSION DU CHAPITRE :**

### **5.1 La modularité de l'innovation :**

Notre travail éclaire la question de l'élaboration des *design rules* qui permettent une modularisation des systèmes et de leur conception. Dans le cas des ordinateurs, ces *design rules* ont été constituées ex-ante par un acteur dominant du secteur, par exemple Intel. Notre recherche met en évidence un autre processus de construction des *design rules* (Baldwin et Clark, 1997). La structure modulaire (ou quasi-modulaire) résulte ici de l'existence d'une pluralité de sociétés logicielles spécialisées qui ont développé des composants spécifiques. Les *design rules* émergent

progressivement lorsque ces entreprises sont amenées à coopérer pour construire une offre globale satisfaisant les besoins d'un client et nécessitant l'association de ces briques. Alors que la dynamique dans l'industrie informatique a été vers une désintégration verticale s'accompagnant d'une modularisation, nous analysons un mouvement inverse de combinaison d'offres spécialisées en une offre globale intégrée (dynamique illustrée par le développement du *component based software engineering* ; Brown et Wallnau, 1998), mais sans intégration des éditeurs de logiciel concernés. Nous mettons alors l'accent sur le travail nécessaire pour rendre les quasi-modules interconnectables, et la conception logicielle qui ne peut se faire qu'en situation de coopération. Nous montrons alors que l'élaboration de *design rules* est répartie entre deux acteurs : les *lead users* qui définissent l'architecture de l'innovation et exigent que plusieurs de ces entreprises coopèrent, et ces entreprises (éditeurs de logiciel experts de leur propre technologie), qui adaptent leur modules et conçoivent les interfaces manquantes afin de les rendre interconnectables. L'ensemble des composants logiciels forme alors une plate-forme logicielle innovante, qui constitue un nouveau type d'offre logicielle (quasi-)modulaire, « système de création automatisée de base de connaissance » à laquelle peuvent se connecter d'autres modules logiciels (moteurs de recherche, crawlers (ou robots d'indexation), systèmes de gestion de contenu ou moteurs d'inférence), sans nécessiter de coopération approfondie entre éditeurs de logiciel.

La définition de modes de coordination entre éditeurs de logiciel est nécessaire, tant que les pseudo-modules de la plate-forme, sont portés par des entreprises différentes. Les éditeurs de logiciel radicalement innovants se focalisant sur leur technologie propre, un rachat ou une fusion entre Temis et Mondeca, semble peu probable à court terme. La définition de modes de coordination entre ces organisations, mais aussi entre ces organisations, le client et l'acteur chargé du développement des interfaces utilisateur, fait donc partie des *design rules* de l'innovation.

## 5.2 L'innovation par les utilisateurs :

Notre recherche met en évidence un client *lead user* dont la compétence est d'abord sur les usages et dont la contribution à la conception de logiciel innovant est sur l'architecture d'une offre complète et le pilotage des premiers projets qui la mettent en œuvre.

Ce résultat dans le domaine du logiciel propriétaire se distingue assez nettement des travaux sur le logiciel libre qui soulignent le rôle des *lead users* dans l'écriture du code des logiciels innovants.

Toutefois, il illustre bien l'importance des sources de connaissance externes, dans le développement d'innovation logicielle, que la littérature a mis en exergue (Grimaldi et Torrisi, 2001 ; Segelod et Jordan, 2004). Dans notre cas, le client sélectionne les « briques » logicielles de différentes entreprises à associer afin de satisfaire son besoin propre et met en place entre ces entreprises des contrats afin d'obtenir la réalisation d'une application intégrée. Cela conduit chacune des entreprises à modifier sa « brique » et à créer les passerelles permettant à l'ensemble de l'application de fonctionner. Les entreprises écrivent également le code nécessaire à l'intégration de cette nouvelle application dans le système informatique du client. L'architecture est conçue par le premier client pour ses besoins spécifiques, et elle est ensuite reprise par les entreprises de logiciel dans leur proposition au second client. Nous montrons alors comment les éditeurs de logiciel étudiés s'approprient l'apport du premier client à la conception du logiciel.

L'évolution constatée entre le premier et le deuxième projet, nous incitent à faire l'hypothèse d'une maturation progressive de l'offre par l'enrichissement des différents projets, et parallèlement d'une maturation de l'interaction entre l'éditeur de logiciel innovant et les premiers utilisateurs. Les premiers projets traversent une phase pendant laquelle des clients, motivés mais ne maîtrisant pas l'innovation, contribuent à faire émerger une offre, en collaborant avec des éditeurs de logiciel qui eux doivent bâtir une offre modulaire ou quasi-modulaire, et l'adapter à un contexte précis, pour la rendre opérationnelle.

En conclusion, la prise en compte des demandes des premiers clients, par les éditeurs de logiciels innovants, peut s'avérer profitable sous certaines réserves que nous explicitons ensuite, et ce pour les raisons que nous détaillons ici :

1. Seuls les clients sont en mesure de concevoir des solutions, quasi-modulaires au départ, combinant des briques logicielles innovantes, du fait du fonctionnement de l'industrie du logiciel et des stratégies d'hyperspécialisation suivies par les éditeurs de logiciel radicalement innovant. La conception de telles innovations est rendue possible par la nature, quasi-modulaire, des offres des éditeurs de logiciel.
2. Les demandes de ces premiers clients permettent de financer la conception d'innovations modulaires ou quasi-modulaires intégrées dans la plate-forme, dont une partie sera générique et donc réexploitable ultérieurement, au travers de projets spécifiques. Sans cela les start-up logicielles ne disposeraient pas des ressources, leur permettant d'envisager des développements en partenariat avec d'autres éditeurs de logiciel innovants.

3. De telles combinaisons sont nécessaires pour qu'émergent de nouvelles plates-formes logicielles, qui pourront ensuite s'imposer sur le marché, plaçant chacun des contributeurs initiaux de cette plate-forme, et plus tard ceux qui proposeront des produits compatibles, dans une position concurrentielle favorable. La nécessité d'identifier et de faire partie aussi rapidement que possible d'une nouvelle plate-forme logicielle, est conditionnée par le fait que, pour un type d'application donné une, voire quelques plate(s)-forme(s) logicielle(s) s'impose(nt) sur le marché, ce qui rend ensuite très difficile l'émergence de plateformes concurrentes.

Toutefois, notre travail fait apparaître un certain nombre de conditions à respecter, pour que de tels projets soient profitables pour les éditeurs de logiciel radicalement innovant. Nous détaillons ces conditions au travers de cinq points :

1. Idéalement, le client doit présenter le profil de client *lead user* et disposer en son sein d'acteurs susceptibles d'assumer les différents rôles que nous avons listé. Etre *lead user* signifie se trouver à la pointe d'une tendance de marché et être en position de retirer un profit substantiel de l'adoption de l'innovation. Une telle évaluation n'a rien d'évident a priori, même si elle semble relativement aisée dans le cas qui nous intéresse, pour ce qui est de la tendance de marché : le traitement de document non structuré a donné lieu à de nombreux travaux de recherche, à de nombreux développements dans les centres de recherche de tout type, à l'émergence d'un certain nombre de progiciels conçus par des start-up, mais aussi parfois de grandes entreprises ; il s'inscrit dans une tendance lourde, celle de l'avènement du web sémantique, annoncée par les experts, et correspond concrètement à une problématique rencontrée par toutes les entreprises, celle de la nécessité d'extraire et d'organiser la connaissance contenue dans les documents non structurés qu'elles possèdent ou auxquelles elles peuvent avoir accès. La question du bénéfice retiré est moins évidente, et dépend des entreprises, toutefois, dans les deux cas étudiés, les gains futurs se situent, au niveau de l'optimisation des processus de traitement de documents, et de la valeur ajoutée de l'offre proposée aux clients.
2. Mais ce n'est pas parce qu'un client possède les caractéristiques d'un *lead user* qu'il sera capable d'orienter, de superviser et de gérer le projet, et encore moins de réaliser des développements personnalisés par lui-même. Les éditeurs de logiciel sont amenés à prendre en compte le niveau de compétences de leur client, qu'ils risquent par ailleurs de surestimer. Cela peut les conduire à mettre en place des dispositifs permettant un transfert

de compétences progressif chez ces clients, mais aussi à des sociétés de service informatique, futures intégratrices de leur innovation. Les fonctions de directeurs et de gestionnaire de projet, au sens « light weight product manager », ne pouvant qu'être difficilement assumée par les clients ou par des SSII, au cours des premiers projets, les éditeurs de logiciel innovants doivent anticiper la nécessité d'endosser ces rôles.

3. Le client *lead user* doit à assumer tout ou partie des trois rôles suivants :
  - a. Architecte innovation : rôle assumé par un « *lead user* architecte innovation » situé chez le client, qui imagine, oriente et supervise le projet, qui joue le rôle de « heavy weight project manager » (Clark et Wheelwright, 1992) ou « directeur de projet » (Giard et Midler, 1993).
  - b. Concepteur/réalisateur des composants de personnalisation : rôle partagé entre un expert de l'éditeur de logiciel concerné et les « *lead users* experts » situés chez le client, qui analysent les usages des « *lead users* utilisateurs classiques ». Toutefois, le développement informatique des composants, en tant que tel, est effectué par l'éditeur de logiciel.
  - c. Concepteur/réalisateur des interfaces utilisateur, qui permet l'intégration de la solution au système d'information du client : rôle partagé entre les « *lead users* experts », le « *lead user* architecte innovation », le « *lead user* gestionnaire de projet », indispensable maître d'œuvre de la solution, tous trois situés chez le client, et l'acteur en charge de son développement informatique, par exemple une société de services informatiques, possédant des compétences en développement d'interfaces utilisateur et acceptant d'en développer de nouvelles, en représentation des connaissances. Ici aussi, l'analyse des usages des « *lead users* utilisateurs classiques » est requise.
4. La question du pilotage et la coordination du travail des différents éditeurs de logiciel, relatif à la conception de l'innovation et aux premiers projets la mettant en œuvre, est cruciale : elle ne va pas de soi, car elle doit s'effectuer entre organisations faiblement couplées, et qui peuvent le rester pendant un certain temps. Cette coordination doit être pensée et organisée pour une autre raison : elle porte sur la constitution d'une solution à partir de ce que nous appelons des quasi-modules, susceptible de nécessiter des développements concourants de passerelles logicielles et un travail de personnalisation. De plus, l'interdépendance de certaines personnalisations, menées par des organisations différentes, accroît le besoin de coordination. Ainsi, la coordination des premiers projets

au cours desquels émerge la nouvelle solution logicielle, la coordination nécessaire à la conception et au développement des nouveaux composants logiciels requis, et la coordination des projets qui s'appuieront sur ces composants développés au cours des premiers projets, constituent trois niveaux de coordination différents qui doivent être pensés par les éditeurs de logiciel. Le ou les premiers projets étant exploratoires, la coordination en revient principalement au « *lead user* architecte innovation » qui possède une vision globale du projet et des compétences transversales, qui favorisent sa capacité à interagir avec les différents acteurs. Il est assisté dans cette tâche par un « *lead user* gestionnaire de projet » aux compétences moins étendues, mais qui se charge de la gestion opérationnelle du projet, sous sa supervision. La coordination des projets, une fois la passerelle logicielle développée, est donc triple, et prise en charge par le client *lead user*, en grande partie :

- a. La coordination globale du projet, qui revient au « *lead user* architecte innovation ».
  - b. La coordination entre la conception de l'interface utilisateur et la conception des différents composants logiciels qu'elle sollicite, assumée par le « *lead user* architecte innovation » assisté du « *lead user* gestionnaire de projet ».
  - c. La coordination inhérente au couplage récurrent de quasi-modules, assumée par les éditeurs des logiciels couplés.
5. Les incertitudes entourant les premiers projets peuvent conduire le client à le remettre en cause à tout moment. C'est pourquoi il est profitable de développer une stratégie visant à permettre l'atteinte d'objectifs intermédiaires, dans un but pédagogique, tout au long du projet, montrant la valeur ajoutée qu'il est susceptible d'apporter au client. Pour atteindre cet objectif difficile, le concours du « *lead user* architecte innovation » chez le client est nécessaire, en ce qu'il connaît les enjeux internes de son entreprise. Il s'agit là d'un autre aspect de son rôle dans le processus d'innovation.

### 5.3 L'apprentissage :

Même s'il existe un client *lead user* capable d'assumer les quatre rôles évoqués, l'apprentissage inter-projet s'effectue chez les éditeurs de logiciel, que ce soit au niveau des logiciels eux-mêmes (modules ou pseudo-modules), au niveau de la plate-forme logicielle qui émerge des projets, de la méthodologie de conduite de projet adaptée à la mise en œuvre de l'innovation (dont la

« démarche OntoPop »), de l'évaluation des compétences des acteurs, et de l'identification des rôles.

En effet, si les différents logiciels cœur n'ont que peu évolué suite aux projets PressPro et Exinis, Temis et Exinis ont capitalisé sur les personnalisations qu'ils ont créées. Temis a capitalisé sur le développement de cartouches de connaissance adaptées à des types de textes non encore pris en compte, tels que les articles de la presse « people », ce qui lui a permis ensuite de réussir un projet d'annotation de documents photographiques à l'AFP. De même, Exinis a capitalisé sur l'organisation de la connaissance que recèlent les textes juridiques français, ce qui lui a permis de se lancer dans un projet du même type en Grande-Bretagne. La capitalisation porte à la fois sur des éléments tangibles, tels que des lignes de code qui peuvent être réutilisées, mais aussi sur des éléments de savoir-faire, comme la façon d'appréhender un projet traitant de textes d'un certain type. 4D Concept a aussi capitalisé sur le projet Exinis, et développé des compétences qui lui ont permis d'intégrer avec succès les projets ultérieurs mettant en œuvre le couplage IDE/ITM. De même, la capitalisation relative au couplage entre le système de représentation des connaissances ITM, et le moteur d'extraction d'information IDE, concerne exclusivement Temis et Mondeca. Mondeca surtout, qui emploie la chef de projet couplage des deux projets étudiés, ainsi que l'équipe informatique qui a réalisé les composants de la passerelle OntoPop.

La succession de projets dans des champs forts différents a fourni à Temis, Mondeca et dans une moindre mesure à 4D Concept, des éléments permettant une étude comparative des rôles joués par les acteurs de ces projets et de leurs compétences respectives. Cela place ces fournisseurs dans une situation plus favorable que leurs clients pour évaluer les compétences et identifier les rôles de ces acteurs. De même, cette évaluation et cette identification est facilitée par leur niveau de compétences en *text mining* ou en gestion des connaissances, bien supérieur à celles de leurs clients. L'expérience de projets variés et le niveau de compétences relatif aux technologies employées favorise donc la capitalisation des apprentissages chez les fournisseurs.

La capitalisation des apprentissages chez le client a commencé à être observée chez Exinis, six mois après la mise en production du système. Tout d'abord, les utilisateurs experts sont parvenus à un premier niveau d'appropriation de la solution, au travers de la réalisation autonome de tests de non régression, suite à la mise à jour des lexiques du moteur d'extraction, IDE. Par contre, toujours à cette date, le développement autonome de nouvelles cartouches de connaissance, grâce au *toolkit* de développement logiciel, ou de Règles d'Acquisition de Connaissances avec le logiciel

éditeur de RAC, n'était toujours pas opérationnel, ni envisagé à court terme, comme nous l'avons vu. Le niveau des compétences de ces utilisateurs experts n'avait pas encore assez évolué pour leur permettre de réaliser ces deux tâches. Par ailleurs, trois nouveaux projets reposant sur la plate-forme étudiée, ont débuté chez Exinis, courant 2008. Nous pouvons donc conclure que cette capitalisation est limitée, et concerne avant tout la compréhension de projets reposant sur le couplage IDE/ITM, et le savoir-faire lié à leur gestion par l'architecte innovation, en premier lieu, et par le chef de projet intégration informatique (gestionnaire de projet), ensuite. Cela rend illusoire la perspective de l'émergence d'une communauté d'utilisateurs échangeant au sujet du développement de cartouches de connaissance ou de Règles d'Acquisition de Connaissances.

Or, la capitalisation des apprentissages chez les fournisseurs, n'est pas soulignée par la littérature sur l'innovation par les utilisateurs, qui insiste au contraire sur l'apprentissage réalisé par les utilisateurs chez le client (Tyre et von Hippel, 1997) et au sein des communautés d'utilisateurs (von Hippel, 2005 ; von Hippel et von Krogh, 2003 ; Franke et von Hippel, 2003). Il nous paraît important de souligner, au contraire, une capitalisation chez les concepteurs des logiciels cœur, Temis et Mondeca, qui fondent la plateforme modulaire étudiée, ce qui corrobore les conclusions avancées par la littérature portant sur l'apprentissage organisationnel (Charue-Duboc, 2005). Cette littérature insiste sur l'importance de l'apprentissage inter-projet (Nonaka, 1994 ; Nobeoka et Cusumano, 1995), et met en évidence des processus de construction des connaissances dans le cadre de la conception de nouveaux produits au travers de projets successifs (Hatchuel, Weil et Le Masson, 2000), qui correspondent effectivement à ce que nous avons observé au cours de notre recherche, principalement chez les fournisseurs.

# **Conclusion**

---

## 1. LES APPORTS DE LA RECHERCHE :

### 1.1 Introduction :

L'intérêt principal de notre recherche vient de l'analyse en temps réel du processus d'innovation, de la conception de l'architecture produit à la mise en œuvre d'une innovation logicielle radicale, portée par deux éditeurs de logiciel. L'intérêt en est d'autant plus grand que cette innovation se situe dans un champ applicatif lui-même nouveau, caractérisé par un petit nombre d'acteurs, des start-up technologiques essentiellement. Nous avons eu l'opportunité rare, d'observer une situation emblématique, l'émergence d'un marché, en partant d'une situation où les technologies employées par l'innovation sont maîtrisées par un petit nombre de personnes généralement issues de la recherche, où les applications, en l'occurrence, celles du *text mining* et de la représentation des connaissances, sont peu connues des entreprises, et où il n'existe pas encore d'acteur(s) majeur(s), ni de logiciel plate-forme dominant le marché. Notre travail de recherche a permis de caractériser un processus d'innovation original qui est celui d'une innovation quasi-modulaire multi-acteurs : ce processus englobe un processus de modularisation et un processus d'apprentissage relatif à la mise en œuvre de cette innovation dans les entreprises clientes. Le processus de modularisation étudié se distingue clairement de ceux décrits par la littérature. Nous décrivons ce processus, qui pose des questions qui ont trait au pilotage et à la coordination des tâches, à la nature et aux compétences des acteurs, et nous proposons des catégories nouvelles en lien avec celui-ci. Ce processus d'innovation accorde, voire impose, un rôle particulier aux premiers clients, rôle qu'ils ne sont pas forcément en mesure d'assumer. Nous apportons donc une contribution à la littérature sur l'innovation par les utilisateurs, en soulignant dans un premier temps les différents types d'interactions et de collaborations dans lesquels le client est impliqué : nous constatons ensuite que les deux premiers clients de l'innovation radicale étudiée, présentent des caractéristiques de *lead users*, telles que les décrit la littérature. Mais nous constatons aussi qu'il existe des contributions multiples de la part de différents acteurs ou groupes d'acteurs situés chez ces clients : en conséquence, nous sommes amenés à enrichir la notion de *lead user*, en la rendant multidimensionnelle. Nous montrons aussi que la contribution de ces premiers clients, si elle est déterminante, n'en demeure pas moins limitée, pour des raisons dues, principalement, à leurs compétences et à l'apprentissage réalisé par les éditeurs de logiciel au fil des projets.

## 1.2 Apports dans le champ de la modularité :

Notre travail de recherche décrit un processus de création de produit doté d'une architecture que nous appelons « quasi-modulaire », qui se distingue des processus de dé-intégration présentés par la littérature. Il s'agit bien là d'un processus de modularisation particulièrement original. Partant, de composants (logiciels) existants, déjà commercialisés de manière autonome, ce processus requiert des collaborations entre les différents éditeurs de logiciel concernés et leur client. Nous qualifions de « quasi-modulaire », l'architecture initiale de l'innovation logicielle imaginée à partir de composants indépendants, dont l'interfaçage n'est pas défini au début du processus d'innovation. Ce terme, « quasi-modulaire », se distingue de celui retenu par la littérature sur la modularité (Baldwin et Clark, 1997), « pré-modulaire », qui qualifie une architecture de produit intégré que l'on cherche à dé-intégrer. Dès son démarrage, le processus d'innovation, qui englobe le processus de modularisation, voit la mise en place de collaborations interorganisationnelles et intra-organisationnelles, impliquant les éditeurs de logiciel concernés, mais aussi le client, et dans le cas du projet Exinis, une société de services informatiques. Les collaborations intra-organisationnelles requises ne bouleversent pas, en tant que telles, le fonctionnement des organisations concernées. Il faut toutefois noter, sur ce point, que des changements ont tout de même affecté l'organisation interne de certaines entreprises : Mondeca a du recruter une chercheuse spécialiste de la gestion informatisée des connaissances, pour travailler au couplage des logiciels IDE/ITM. De même, le client Exinis, a mis en place une cellule composée de plusieurs personnes dédiée au projet. Mais, les collaborations interorganisationnelles présentent un caractère exploratoire qui constitue une des caractéristiques propres au processus d'innovation étudié. Le caractère exploratoire de ce processus lors du tout premier projet notamment, est multidimensionnel, car il concerne les tâches, les collaborations interorganisationnelles nécessaires à leurs réalisations, et la coordination de ces tâches :

1. Des tâches ont été définies par le cahier des charges, mais elles sont en partie nouvelles, pour chacun des acteurs. Il y a donc tout d'abord la découverte de certaines des tâches à réaliser, notamment :
  - le couplage IDE/ITM, tâche la plus nouvelle et la plus complexe,
  - l'application documentaire, qui constitue l'interface utilisateur du système, tâche nouvelle pour chacun des acteurs du projet PressPro,

- la définition d'une cartouche de connaissance par le linguiste de Temis, dans un domaine où cela n'avait jamais été encore fait (la presse grand public « people »), ce qui pose des problèmes techniques particuliers.
2. La dimension exploratoire du processus concerne aussi les collaborations à mettre en œuvre dans le cadre de ce processus : il faut définir les modalités de collaboration impliquant les acteurs de différentes organisations. Ces organisations sont indépendantes, et n'avaient pas collaboré avant le projet PressPro. Dans le cadre du projet PressPro, aucun acteur n'a été en mesure d'exercer une autorité hiérarchique, que ce soit le client par manque de vision globale du projet et de compétences, ou l'un des éditeurs de logiciel impliqués, par manque de compétences relatives à la direction d'un projet multi-acteurs, par manque de ressources à imputer à cette tâche, mais aussi par manque de compétences relatives aux technologies des autres éditeurs. Les acteurs mettent donc en place une série de collaborations, dans un contexte de couplage organisationnel faible.
  3. Une autre dimension exploratoire est constituée par la coordination entre les tâches : le premier projet permet de mettre en évidence que le besoin de coordination n'avait pas été anticipé correctement, qu'il était réel et qu'il était triple. Le besoin de coordination étant défini, il reste à déterminer quels sont acteurs ou groupes d'acteurs susceptibles de les assumer. Les enseignements du premier projet conjugués à ceux du second, ont permis de mettre en évidence les caractéristiques de ces acteurs ou groupes d'acteurs. Il apparaît que le client a ici un rôle particulier à jouer, sur lequel nous revenons dans le point suivant relatif à nos apports dans le champ de l'innovation par les utilisateurs.

Les collaborations à l'œuvre dans les projets étudiés, visent d'abord à la définition de la nature des composants logiciels constitutifs de la solution recherchée. Nous proposons une classification de ces composants en quatre catégories, selon qu'ils sont modulaires ou « pseudo-modulaires » d'une part, génériques ou ad hoc d'autre part. En effet, parmi ces composants, certains, que nous appelons « pseudo-modules », sont des composants non modulaires, car leur développement reste dépendant de celui d'autres « pseudo-modules ». Ainsi, peut-on considérer que le processus de modularisation n'est pas, au moment où nous achevons cette étude, complètement arrivé à son terme : si ce que nous avons appelé la « plate-forme de création automatisée de base de connaissance », qui regroupe les logiciels IDE, ITM et la passerelle OntoPop, peut effectivement être considérée comme un module, sur lequel peuvent venir se connecter des logiciels tels que *crawler*, moteur de recherche ou système de gestion de contenu, à l'intérieur de cette plate-forme,

---

les pseudo-modules doivent être développés les uns en fonction des autres par des sociétés différentes : il en va ainsi des RAC (règles d'acquisition des connaissances), de la cartouche de connaissance personnalisé, et de l'ontologie adaptée au domaine étudié.

Cette étape qui voit émerger une plate-forme composée de modules et de pseudo-modules, n'est peut-être que transitoire, mais se justifie par le fait que chaque éditeur de logiciel développe une activité commerciale qui est d'abord fondée sur la vente de ces logiciels cœur, ou d'une suite logicielle les combinant (exemple de la suite Luxid de Temis). Il ne saurait alors être question pour eux de focaliser leurs efforts techniques et commerciaux sur la nouvelle offre couplée, même si celle-ci fonde un nouveau type de plate-forme applicative dans le domaine du *document processing*. Bien sûr, chaque éditeur de logiciel connaît l'importance d'une participation à une nouvelle plate-forme applicative, en ce que celle-ci est susceptible de devenir un standard. Mais ces sociétés sont d'abord préoccupées d'asseoir leur technologie et de la déployer dans un maximum de grands comptes pour, dans un premier temps, devenir une référence dans leur champ technologique propre. Soumis à l'influence de ces deux attracteurs, les éditeurs de logiciel, ne veulent pas prendre le risque de consacrer l'essentiel de leur activité à un marché encore balbutiant et incertain, dans un contexte de rareté des ressources humaines et financières. Et ce d'autant qu'une telle éventualité les contraindrait à des collaborations poussées avec d'autres start-up technologiques, dont elles perçoivent la fragilité en tant qu'elle est comparable à la leur. Ainsi, le processus de modularisation peut-il se stabiliser à un stade, où cette dernière n'est pas achevée. Ce stade peut être dépassé de différentes manières, pour aboutir à un produit dont chaque composant est un module : cela pourrait être le cas, si par exemple Temis rachetait Mondeca, ce à quoi aspirent certains acteurs des projets étudiés chez Temis. Cela pourrait aussi survenir si les deux éditeurs de logiciel étaient rachetés par une même entreprise, comme IBM ou Microsoft. Enfin, cela pourrait se produire, si l'un des éditeurs de logiciel développait ou acquérait des compétences possédées par l'autre : ainsi, la suite logicielle Luxid de Temis englobe déjà un système de gestion de base de connaissance. Dans l'attente d'une éventualité de ce type, nous constatons donc que l'architecture quasi-modulaire de l'innovation, peut rester à un stade où elle n'encastre pas complètement la coordination entre les organisations qui conçoivent et développent ses composants.

Nos apports dans le champ de la littérature sur la modularité concernent donc la typologie des différents composants logiciels d'une plate-forme logicielle radicalement innovante « quasi-modulaire », et la caractérisation d'un processus de modularisation original.

Les deux projets que nous avons étudiés ont permis d'apporter une grande partie des réponses relatives au rôle des acteurs du processus d'innovation, dont les premiers clients. Ces projets ont fait émerger, après une phase d'exploration, la nécessité de recourir à trois types de coordination, global, lié au couplage IDE/ITM, lié au développement de l'interface utilisateur (ou application documentaire). Ces coordinations ne peuvent être assumées par le même acteur. Le client est le mieux placé dans le premier cas, un des deux éditeurs des logiciels couplés dans le deuxième, un spécialiste de l'intégration informatique ayant l'expérience de l'intégration de système informatique et du développement d'interfaces utilisateur dans le troisième. Les collaborations entre éditeurs, et donc la coordination liée au couplage IDE/ITM, sont largement facilitées par la proximité intellectuelle et culturelle, entre les employés de Temis et de Mondeca : il s'agit d'informaticiens possédant pour la plupart une formation en sciences du langage et/ou gestion des connaissances, issus pour certains du même cursus universitaire. Cette situation n'est pas surprenante dans le sens où, Temis et Mondeca appartiennent toutes deux à un même champ et emploient un certain nombre de diplômés récemment issus de la recherche. La difficulté est ici plutôt liée au fait qu'il s'agit d'effectuer des tâches nouvelles, impliquant la coordination de trois tâches (définition des patrons d'extraction, définition de l'ontologie, et enfin définition des Règles d'Acquisition des Connaissances).

Dans le cas du client, la proximité culturelle avec les éditeurs de logiciel n'est pas la même, la question des compétences de certains acteurs, dont nous avons défini les rôles en conclusion de la discussion relative à l'innovation par les utilisateurs, est cruciale. Sur ce point, la difficulté pour les éditeurs de logiciel impliqués vient de ce qu'ils ne sont pas en mesure de maîtriser cet aspect.

En synthèse, le processus d'innovation inclut le processus de création d'une architecture modulaire au travers de la création de composants logiciels nouveaux, la définition d'une méthodologie adaptée à la mise en œuvre de l'innovation ainsi définie, et plus largement l'apprentissage relatif aux rôles qui doivent être assumés par les différents acteurs de cette mise en œuvre, ainsi qu'à la nécessaire personnalisation de certains composants. Les acteurs de ce processus auront donc été PressPro, Temis et Mondeca dans un premier temps, puis Exinis et à nouveau Temis et Mondeca. Les rôles tenus par Xyleme dans le cadre du premier projet, et par 4D Concept dans celui du second, n'ont pas contribué à infléchir le processus d'innovation : dans le premier cas, le logiciel de Xyleme s'est *pluggé* en tant que module, sur la plate-forme formée par le couplage IDE/ITM, tandis que dans le second, 4D Concept a développé l'interface utilisateur, ce qui ne nécessite, finalement, pas une connaissance parfaite des technologies employées.

### 1.3 Apports dans le champ de l'innovation par les utilisateurs :

Notre recherche conduit à réexaminer la notion de *lead user* telle qu'elle est définie par la littérature sur l'innovation par les utilisateurs. Nous constatons au terme de cette étude, que cette notion est effectivement riche et précieuse lorsqu'il s'agit d'étudier les processus d'innovation radicale dans le logiciel. Mais nous estimons aussi que les résultats de notre travail conduisent à enrichir cette notion en la rendant multidimensionnelle, tout en en limitant la portée.

La littérature sur l'innovation par les utilisateurs décrit, particulièrement dans le cas du logiciel, un processus d'innovation au cours duquel des *lead users* s'emparent de modules logiciels personnalisables, et définissent avec une précision qu'aucun fournisseur ne pourrait égaler, une solution parfaitement adaptée à leur besoin, besoin qui deviendra celui de la masse du marché à terme. Ce n'est pas ce que nous observons, même si nous constatons le rôle crucial des premiers clients, dont les utilisateurs finaux dans le processus d'innovation. En effet, nous observons que des clients présentant les caractéristiques de *lead users* sont indispensables à ce processus d'innovation logicielle radicale. Ici, c'est d'abord dans la conception de l'architecture produit nouvelle que le client joue un rôle essentiel. Les éditeurs de logiciels concernés ne sont pas dans une situation aussi favorable pour cela, principalement du fait de leur focalisation technologique. Cette contribution est donc très forte, puisqu'elle dépasse même celle décrite par la littérature sur l'innovation par les utilisateurs portant sur le logiciel : celle-ci insiste surtout sur la capacité des *lead users* à concevoir des logiciels à partir de modules logiciels personnalisables, les *user toolkits*, connectés à des modules logiciels génériques.

Ensuite, le premier projet montre toutes les difficultés que ce client, pourtant *lead user*, rencontre pour participer efficacement au processus d'innovation : problèmes liés à la coordination et au pilotage global du projet, ainsi qu'au développement des interfaces utilisateur, qui conduisent à un échec. Ainsi, il apparaît qu'il ne suffit pas de ressentir des besoins à la pointe d'une tendance de marchés (le *document processing*) et d'espérer retirer des bénéfices élevés de la satisfaction de ces besoins, pour être un *lead user* efficace dans le processus d'innovation logicielle radicale.

---

L'étude des deux projets montre que le client, peut et doit jouer un rôle par l'intermédiaire d'acteurs ou de groupes d'acteurs différents, qui doivent disposer de certaines compétences :

- Le « *lead user* architecte innovation »,
- Le « *lead user* gestionnaire de projet»,
- Le ou les « *lead user(s)* expert(s) ».

Les « *lead user* utilisateurs classiques » sont simplement des utilisateurs sans compétences particulières dans les technologies qui fondent l'innovation, mais dont la participation est nécessaire pour cerner l'ensemble des usages et effectuer des tests. Il faut donc savoir les intégrer au projet, par le biais des « *lead users* experts », notamment. Les « *lead users* » dont le profil se rapproche le plus de ceux décrits par la littérature (von Hippel, 1986) sont justement ceux que nous avons nommés « *lead users* experts » : ce sont de véritables utilisateurs du futur système, qui possèdent des compétences transversales particulières, et qui, de ce fait, ressentent des besoins à la pointe d'une tendance de marché, et comprennent les bénéfices qu'ils pourront retirer de l'innovation. Ils sont indispensables pour faire le lien entre l'utilisation concrète par le client et les technologies nouvelles employées. Mais leur position d'utilisateurs, ne les place pas forcément dans des conditions idéales, ni ne les dote forcément des compétences nécessaires pour endosser les différents rôles des quatre types de *lead user* que nous avons définis : ainsi, ils ne peuvent pas forcément assumer le rôle de « *lead user* architecte innovation », rôle qui requiert une vision globale des technologies et des acteurs qui les portent, une capacité à penser le projet et ses prolongements à moyen voire long terme, et la capacité à interagir avec la direction générale de sa propre entreprise afin de répondre aux doutes qu'elle peut être amenée à ressentir de manière récurrente vis-à-vis de l'innovation. Il en va de même pour les compétences de « *lead user* gestionnaire de projet», tout utilisateur d'un système informatique ne possédant pas nécessairement de compétences en gestion de projet logiciel.

Ainsi, l'existence chez le client d'un ou plusieurs *lead user(s)* « monodimensionnel(s) » (le ou les « *lead user(s)* expert(s) ») n'apparaît pas suffisant pour que le processus d'innovation puisse aller à son terme. De plus, le fait que le client recèle en son sein les quatre types de *lead users* que nous avons évoqués, ne suffit pas non plus pour que celui-ci puisse réaliser l'ensemble du processus d'innovation.

A cela, nous voyons deux raisons majeures : d'abord, le couplage de logiciels requiert le développement de compétences nouvelles à partir des compétences relatives aux technologies employées par chacun des logiciels couplés : il faut donc développer des compétences nouvelles à partir de compétences possédées par quelques start-up logicielles. En conséquence, ce sont bien les éditeurs de logiciel concernés qui sont les mieux placés pour cela. Par ailleurs, un certain nombre de connaissances sont capitalisées par les éditeurs de logiciel, au fur et à mesure des projets : cet apprentissage concerne la méthodologie à mettre en œuvre, la coordination des tâches, la personnalisation des logiciels. Le client pour lequel un tel projet est mis en œuvre, ne peut posséder ce type de connaissance, n'ayant pas pu bénéficier d'un tel effet d'apprentissage. Il se peut toutefois qu'à terme, un client chez lequel une succession de projets aura été déployée, développe une expertise propre, mais ce n'est pas encore le cas au moment où nous achevons ce travail de recherche. En juin 2008, les outils de personnalisation destinés aux utilisateurs finaux, étaient encore d'un usage délicat pour les utilisateurs finaux, voire en cours de développement (éditeur de RAC, environnement de développement STDK), et n'étaient donc pas encore utilisés par le client Exinis, malgré sa volonté de « prendre le contrôle » de la solution. Cela montre la difficulté d'une telle démarche, et indique qu'on ne peut s'attendre à ce que des utilisateurs développent leur propre personnalisation du « système de création automatisée de base de connaissance », à court terme. Quand cette appropriation de la dimension « personnalisation » du logiciel sera réalisée par certains clients, alors des échanges entre utilisateurs pourront prendre place, créant ainsi les conditions d'établissement d'une communauté d'utilisateurs.

## **2. LIMITES DE LA RECHERCHE :**

Nous soulignons deux types de limitations quant au matériau empirique analysé. L'une a trait à la fenêtre temporelle dans laquelle nous avons étudié ce processus d'innovation, qui est nécessairement bornée. La seconde a trait au matériau auquel nous avons pu avoir accès du fait de la méthodologie que nous avons choisie.

### **Des processus longs :**

La durée des projets et donc du processus d'innovation a largement excédé les délais initialement fixés, ce qui a conduit à allonger la durée de notre travail de recherche. De plus au terme de celui-ci, de nouvelles perspectives apparaissent. Le couplage IDE/ITM requiert une collaboration interorganisationnelle poussée, mais cette situation peut évoluer de différentes manières, les

---

organisations concernées étant susceptibles de se transformer au travers de fusions, de rachats ou du développement d'activités nouvelles, comme nous l'avons expliqué au point 1.2 de cette conclusion.

Par ailleurs, le processus d'innovation est susceptible de se prolonger jusqu'à ce que des outils de développement conviviaux permettent aux utilisateurs de développer leur propre cartouche de connaissance et de réaliser leurs propres règles d'acquisition de connaissance. Cela nécessite bien sûr de nouveaux développements informatiques, mais cela requiert aussi une dissémination des connaissances relatives à l'extraction terminologique et à la gestion des connaissances, ce qui dépasse le simple rôle d'éditeur de logiciel.

Ensuite, d'un point de vue plus global, la plate-forme fondée sur le triptyque IDE/passerelle OntoPop/ITM a connu d'autres applications, à l'AFP, chez Thomson Scientific, et dans un autre grand groupe de presse français. Cela démontre que l'innovation étudiée peut être commercialisée avec succès dans le secteur de la presse et de l'édition, et qu'elle ne constitue pas un développement spécifique. Mais, le nombre de ces applications est encore limité. Dans ces conditions, il est probable que l'innovation évolue encore au fil des projets la mettant en œuvre. Enfin, le standard UIMA mis en place par IBM, peut avoir un impact sur les outils employés dans le domaine de la « gestion d'information non structurée », et donc sur la plate-forme que nous étudions, en facilitant les couplages de logiciels, ce qui peut conduire à l'apparition de nouveaux concurrents mais aussi contribuer à ouvrir le marché.

### **Des entretiens en nombre limités, en début de projet, avec les clients.**

Le choix d'une analyse processuelle nous apparaît judicieux, car la perception par les acteurs du rôle et des caractéristiques de chacun, tout comme celle de l'organisation et des difficultés du projet, a évolué régulièrement au fil du temps, parfois d'une manière radicale. De même, la confrontation de points de vue différents sur une même situation, a permis d'affiner la compréhension que nous en avons eue. Mais, comme nous l'avons expliqué, notre position de chercheur a dû être construite progressivement, et de ce fait nous n'avons pas pu accéder immédiatement aux différents acteurs. Dans la perspective d'une analyse processuelle, destinée, entre autres, à limiter les biais de rationalisation a posteriori et de décontextualisation, cela constitue une des limites de ce travail. Mais nous estimons que ces limites devaient être acceptées

---

afin de pouvoir débiter notre travail de recherche, et de parvenir à observer le processus d'innovation depuis son origine.

### 3. POURSUITE DES RECHERCHES, PISTES DE RECHERCHES FUTURES :

Les domaines du *text mining* et de la gestion des connaissances que nous avons explorés lors de cette recherche, et auxquels nous nous sommes familiarisé, paraissent constituer à l'heure actuelle, un terrain d'étude pertinent pour l'analyse des processus de modularisation et d'innovation multi-acteurs. Nous estimons que les évolutions que nous voyons poindre méritent que des recherches soient poursuivies dans ce champ. En effet, cela permettrait de compléter notre travail sur le processus d'émergence d'une architecture modulaire, en mettant en lumière d'autres phénomènes qui interfèrent avec ce processus, comme l'émergence de standard(s) et l'apparition d'acteur(s) dominant(s).

Ainsi, la démocratisation de standards tels que le standard UIMA, promu par IBM, et l'éventuelle apparition de standards concurrents vont jouer un rôle dans l'évolution de ce secteur. De même d'autres plates-formes apparaîtront, étant donné le nombre coirssant de start-up technologiques présentes dans les domaines évoqués. Enfin, des logiciels Open Source vont se répandre à mesure que les applications de *text mining* se démocratiseront. Ces facteurs vont affecter le développement des futures innovations dans ce secteur. De même, le développement par Temis et Mondeca de *user toolkits* opérationnels, doit permettre le développement par les clients de solutions personnalisées, et, ultérieurement, favoriser l'apparition de communautés d'utilisateurs susceptibles d'échanger à propos de leur expérience à ce sujet. Le rôle du client dans les futurs processus d'innovation en sera forcément modifié. La connexion de nouveaux types de logiciels à la plate-forme étudiée, peut aussi ouvrir de nouveaux champs d'application, et en retour conduire à faire évoluer celle-ci. Au-delà des problématiques liées à la modularisation des architectures produit et au rôle des utilisateurs dans les processus d'innovation, nous touchons là, à la dynamique de l'industrie logicielle, un objet d'étude dont l'intérêt nous parait loin d'être épuisé.

# **Table des tableaux**

---

Tableau 1.	Tableau de comparaison entre les éléments ontologiques et les étiquettes sémantiques _____	190
Tableau 2.	Opérations de consolidation réalisées par OntoPop en fonction des deux axes traités. _____	201
Tableau 3.	Les étapes de la « méthodologie OntoPop » _____	208
Tableau 4.	Typologie des composants logiciels d'une innovation logicielle multi-acteurs _____	257
Tableau 5.	Principaux acteurs des projets PressPro et Exinis (fonctions dans l'entreprise, rôle dans le ou les projets) _____	260
Tableau 6.	Rôles joués par les acteurs situés chez le client et les autres acteurs dans le processus d'innovation. _____	300
Tableau 7.	Rôles joués par les acteurs situés chez le client et par les autres acteurs dans la structuration des projets. _____	301

# **Table des figures**

Figure 1.	Rentabilité des éditeurs de logiciel _____	28
Figure 2.	Rentabilité des éditeurs de logiciel dépassant le milliard de dollars de chiffre d'affaires annuel _____	29
Figure 3.	Le modèle d'interaction _____	93
Figure 4.	Un cadre intégrateur pour quatre de marche de recherché en sciences de gestion _____	104
Figure 5.	Méthodes mises en œuvre dans la recherche sur la conception _____	105
Figure 6.	Fossé entre la représentation textuelle et la représentation sémantique _____	145
Figure 7.	Extrait de l'arbre conceptuel produit par l'analyse linguistique de l'article « la tribu Coppola » publié dans le magazine Elle. _____	149
Figure 8.	Transformation de deux sous-arbres conceptuels en un graphe conceptuel où la même entité nommée « Coppola » fait le lien entre les deux propositions. _____	151
Figure 9.	Fossé entre la représentation textuelle et la représentation sémantique _____	188
Figure 10.	La passerelle logicielle OntoPop _____	188
Figure 11.	Contexte d'une étiquette sémantique dans un arbre conceptuel _____	191
Figure 12.	Exemple d'une étiquette sémantique en langage OPAL. _____	193
Figure 13.	Cycle de vie des RTO dans OntoPop. _____	197
Figure 14.	Processus de consolidation des informations extraites _____	200
Figure 15.	Opérations de consolidation des informations extraites _____	203
Figure 16.	Sources de données pour le linguiste et l'ontologue dans la phase d'étude. _____	211
Figure 17.	Comparaison de la couverture de l'ontologie du domaine et de la couverture de l'analyse linguistique _____	212
Figure 18.	Extrait d'un document de spécifications détaillées des arbres conceptuels délivré par un linguiste de Temis pour une application d'intelligence économique, l'exemple est « isManagerOf » _____	215
Figure 19.	Echanges entre les intervenants pendant la phase de structuration _____	216
Figure 20.	Analyse des problèmes rencontrés par le responsable de couplage _____	219
Figure 21.	Phase de validation _____	221

---

Figure 22.	Phase de mise en service _____	222
Figure 23.	Modules et Composants non modulaires formant la solution PressPro _____	251
Figure 24.	La passerelle logicielle OntoPop _____	254
Figure 25.	Exemples de connexions de différents types de logiciel à la plate-forme de création automatisée de base de connaissance. _____	255
Figure 26.	Modules et pseudo-modules formant la solution PressPro (composants génériques et composants ad-hoc (ou paramétrages). _____	257
Figure 27.	Exemples d’annotations générées aux différents niveaux morphologique, syntaxique et sémantique d’une analyse linguistique _____	355
Figure 28.	Exemple d’une annotation sémantique orchestrée par une ontologie de référence _____	356
Figure 29.	Extrait d’une taxonomie sur la représentation simplifiée de la faune _____	360
Figure 30.	Les différentes relations qui composent un thesaurus. _____	360
Figure 31.	Exemple d’une ontologie dans le domaine de la presse « People » _____	362
Figure 32.	Extrait de l’article « Le Clan Coppola » paru dans le magazine ELLE _____	364

# **Bibliographie**

Abernathy W. J. et Clark K. B. (1985), "Innovation: Mapping the Winds of Creative Destruction", *Research Policy* 14, 3-22.

Akrich M. L., Callon M., et Latour B. (1988), "A quoi tient le succès des innovations ?", *Gérer et Comprendre*, Annales des Mines, 4-17 (n°11); 14-29 (n°12).

Allen R. C. (1983), "Collective Invention." *Journal of Economic Behavior and Organization* 4, no. 1: 1-24.

Allen T. (1984), *Managing the Flow of Technology*. MIT Press, Cambridge, MA.

Alter N. (2000), *L'innovation ordinaire*, Paris, PUF.

Amardeilh F. (2007), *Web Sémantique et Informatique Linguistique : propositions méthodologiques et réalisation d'une plateforme logicielle*, Thèse de doctorat en Informatique, Université Paris X – Nanterre.

Amdahl G. M., Blaauw G. A. et Brooks F. P. Jr. (1964), *Architecture of the IBM System/360 [archive]*, IBM Journal of Research and Development, volume 8, n°2, pp. 87-101.

Aoki M., et Takizawa H. (2002), "Modularity: Its Relevance to Industrial Architecture", *The Saint-Gobain Centre for Economic Research 5th Conference*, 7-8 November, Paris.

Armer P. (1980), "SHARE – A Eulogy to Cooperative Effort", *Annals of the History of Computing* Volume 2, Number 2, April / June, 1980 ; pp. 122-129.

Arndt J. (1979), "Toward a Concept of Domesticated Markets," *Journal of Marketing*, 43(Fall): 69-75.

Arthur B. (1989), "Competing technologies, increasing returns and lock-in by historical events", *Economic Journal*, vol. 99, n°3, pp. 116-131.

Arthur B. (1996), "Increasing Returns and the New World of Business", *Harvard Business Review*, July-Aug.; pp. 100-109.

Axelsson B. et Hakansson H. (1984), *Buying to Create Competitive Advantage*, Stockholm: Liber.

Bachelard G. (1970), *La formation de l'esprit scientifique*, Paris, Vrin.

Bagozzi R. P. (1975), "Marketing as Exchange," *Journal of Marketing*, 39, October, 32-39.

Bain & Company, Heap S., Tobkin V. et Seidensticker F.-J. (2004), *The Supersizing of the Software Industry*, disponible à l'URL

- [http://www.bain.com/bainweb/publications/publications\\_detail.asp?id=18017&menu\\_url=publications\\_results.asp](http://www.bain.com/bainweb/publications/publications_detail.asp?id=18017&menu_url=publications_results.asp).
- Baldwin C. Y. et Clark K. (1997), "Managing in the Age of Modularity", *Harvard Business Review*, September-October, pp. 84-93.
- Baldwin C. Y. et Clark, K. (1994), *Modularity-in-design: An Analysis based on the theory of real options*, Working paper 94-039, Harvard Business School, Cambridge, MA.
- Baldwin C. Y., et Clark K. (2003), *Does Code Architecture Mitigate Free Riding in the Open Source Development Model?* Working paper 03-09, Harvard Business School, Cambridge, MA.
- Baldwin C. Y. et Clark K. (2000), *Design rules: the power of modularity*, Cambridge, MA.
- Ballmer S. (2006) *Microsoft Worldwide Partner Conference*, Vancouver, 9-11/07/2006, citations accessibles à l'adresse URL : [http://news.cnet.com/2100-1012\\_3-6094002.html](http://news.cnet.com/2100-1012_3-6094002.html).
- Becheur A. et Gollety M. (2006), « Validation d'une échelle de mesure du lead user », *Revue Française du Marketing*, février, n°206, 1-5.
- Béji-Bécheur A. (1998), *Utilisateur leader et degré de novation : une contribution à l'amélioration de l'analyse des tests de produits nouveaux*, Thèse pour le Doctorat d'Etat en Sciences de Gestion, Université Paris Dauphine.
- Bell C. G. et Newell, A. (1971), "Possibilities for Computer Structures 1971," *Proceedings Fall Joint Computer Conference*. disponible à l'adresse URL : <http://research.microsoft.com/en-us/um/people/gbell/cgb%20files/>
- Ben Mahmoud-Jouini S. et Midler C. (1999) "Compétition par l'innovation et dynamique des systèmes de conception dans les entreprises françaises - Une comparaison de trois secteurs", *Entreprise et Histoire*, N° 23, décembre, pp. 36-62.
- Ben Miled H. et Le Louarn P. (1994), « Analyse comparative de deux échelles de mesure du leadership d'opinion : validité et interprétation », *Recherche et Applications en Marketing*, 9, 4, 23-51.
- Benghozi P.J., Charue-Duboc F. et Midler C. (2000), *Innovation Based Competition & Design Systems Dynamics*, L'Harmattan, Paris.
- Berners-Lee T. (1998), *Weaving the Web*, Harper Eds, San Francisco.
- Berry G. (2008), *Pourquoi et comment le monde devient numérique*, Fayard, Collection Collège de France, Paris.
- Berry, M. (1981), "Le Centre de Recherche en Gestion de l'Ecole Polytechnique" *Economies et Sociétés*, Série Science de Gestion - n°2 - Vol XV (N°10-11-12).
- Blomqvist K. , Tikkanen J. et Möller K. (2002), *Partnering in the High-Velocity Environment - Is the Interaction Theory Valid Approach*, Paper presented at the 18th Annual IMP Conference. Dijon, France, [impgroup.org](http://impgroup.org).

- Boehm B. (1988), "A Spiral Model of Software Development and Enhancement", *IEEE Comput.*, vol 21, pp 61-72.
- Bourigault D., Aussenac-Gilles N. et Charlet J. (2004), « Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas », In *Techniques Informatiques et Structuration de Terminologies*, Pierrel J.-M. et Slodzian, M. (Eds.), Numéro Spécial de la *Revue d'Intelligence Artificielle (RIA)*, 18(1), Hermès, Paris, pp. 87-110.
- Brooks P. F., Jr. (1979), *The Mythical Man-Month*. Addison-Wesley. Reading, MA.
- Brown J. S., et Duguid P. (1991), "Organizational Learning and Communities-of- Practice: Toward a Unified View of Working, Learning, and Innovation." *Organization Science* 2, n°1: 40–57.
- Brown S.L. et Eisenhardt K. M. (1997), "The art of continuous change: linking complexity theory and time-paced evolution in relentlessly shifting organizations", *Administrative Science Quarterly*, vol. 42, n°1, March., pp. 1-34.
- Brown S.L. et Eisenhardt K. M., (1995), "Product development: past research, present findings, and future directions". *Academy of Management Review* 20 (2), 343-378.
- Brusoni S. et Prencipe A. (2001), "Knowledge specialization, organizational coupling and the boundaries of the firm: why do firms know more than they make?" , *Administrative Science Quarterly*, déc. 34 : 388-410.
- Callon M. (1981), «Pour une sociologie des controverses technologiques», *Fundamenta Scientiae*, 12 (4), p.381-399.
- Callon M. (1998), «Actor-network theory, the market test», in Hassar Jljaj (ed.), *Actor Network Theory and after*, Oxford, Blackwell Publishers / The Sociological Review, p.181-195.
- Callon M. et Latour B. (eds), (1985), *Les scientifiques et leurs alliés*, Paris, Pandore.
- Callon M. et Latour B. (1991), *La Science telle qu'elle se fait*, La Découverte Paris.
- Campbell-Kelly P. (2003), *Une Histoire de l'Industrie du Logiciel*, Vuibert, Paris.
- Chamberlin E. H. (1950), "Product Heterogeneity and Public Policy." *American Economic Review*, 40, n° 2: 85–92.
- Chanal V., Lesca H. et Martinet A.-C. (1997), « Vers une ingénierie de la recherche en sciences de gestion ». *Revue Française de Gestion*, n°116, nov.-déc., pp.41-51.
- Chanal V. et Mothe C. (2005), « Concilier innovations d'exploitation et d'exploration. Le cas du secteur automobile », *Revue Française de Gestion*, Vol. 31, n° 154, pp. 173-191.
- Charue-Duboc F. (2007), « Management stratégique de l'innovation technologique », *Mémoire pour l'Habilitation à diriger des recherches*, IAE Université de Lille.

Charue-Duboc F. (2005), "L'apprentissage organisationnel, un concept séduisant, des processus complexes" in Gilbert P., Guérin F., Pigeyre F. [eds.] *Organisations et comportements*, Paris, Dunod , Gestion Sup , pp. 271-300.

Charue-Duboc F. (1997), "Maîtrise d'oeuvre, maîtrise d'ouvrage et direction de projet, pour comprendre l'évolution des projets chez Rhône Poulenc", *Gérer & Comprendre*, N° 49, septembre 1997, pp. 54-64.

Charue-Duboc F. (1995), *Des savoirs en action* (ouvrage collectif, sous la direction de), l'Harmattan, Paris.

Charue-Duboc F. & Midler C. (2001), "Développer les projets et les compétences - Le défi des hiérarchiques dans les métiers de conception", *Gérer & Comprendre*, mars 2001, n°63, p.12-22.

Chesbrough H. & Kusunoki K. (1999/2001), "The Modularity Trap: Innovation, Technology Phase-Shifts, and the resulting limits of virtual organizations", in *Comparative Studies of Technological Evolution Managing Industrial Knowledge: Creation, Transfer and Utilization*, edited by Ikujiro Nonaka and David J. Teece. London: Thousand Oaks, CA: Sage. pp 202-230.

Clark K.B. (1985), "The interaction of design hierarchies and market concepts in technological evolution". *Research Policy*, 14: 235–251.

Clark K. et Fujimoto T., (1991/1995), *Product development performance. Strategy, organization and management in the world auto industry*, Harvard Business School Press, Cambridge, MA.

Clark K. B. et Wheelwright S. C., (1992), *Revolutionizing product development*. Free Press, New York.

Cohen W. et Levinthal D. (1990), "Absorptive Capacity: A new perspective on Learning and Innovation." *Administrative Science Quarterly*, 35, 128-152.

Connell J.L. & Schafer L. (1989), *Structured Rapid Prototyping: An Evolutionary Approach to Software Development*, Yourdon Press, Englewood Cliffs, NJ.

Cook K. S. et R. M. Emerson. (1978). "Power, equity and commitment in exchange networks". *American Sociological Review*, 43: 721 – 739).

Corey E.R. (1956) *The development of market for new materials*, Division of Research, Graduate School of Business Administration, Harvard University, Boston.

Crispino O G. (2003) *Une plate-forme informatique de l'Exploration Contextuelle : modélisation, architecture et réalisation (ContextO) – Application au filtrage sémantique de textes*, Thèse de doctorat, Université Paris IV-Sorbonne, Paris.

Cunningham, H. (1999) *Information Extraction - a User Guide, Research Memo*, 2nde édition, Université de Sheffield, UK, 1999, <http://www.dcs.shef.ac.uk/~hamish/IE/>.

Cusumano, M. (2004), *The Business of Software*, Free Press, New York.

- Cusumano, M. et Noboeka K. (1998), *Thinking Beyond Lean*, Free Press, New York.
- Cusumano, M. (1991), *Japan's software factories a challenge to US management*, Massachusetts Institute of Technology New York Oxford, Oxford University Press.
- Cusumano, M. & Selby, R.W. (1995), *Microsoft secrets*,- Free Press, New York.
- Danneels, E. (2002), The dynamics of product innovation and firm competences. *Strategic Management Journal* 23 (12), 1095-1121.
- David A., Hatchuel A., Laufer R. (éd.), (2001), *Les nouvelles fondations des sciences de gestion : éléments d'épistémologie et de recherche en management*, Vuibert, Paris.
- Desclés J.-P. & Jouis C., (1993) « L'exploration contextuelle : une méthode linguistique et informatique pour l'analyse informatique de textes », in *Proceedings of the International Language Natural (ILN'93)*, Nantes, 427-450.
- Desclés J.-P., Jouis C., Oh-Jeong H.-G. & Reppert D., (1991) « Exploration Contextuelle et sémantique : un système expert qui trouve les valeurs sémantiques des temps de l'Indicatif dans un texte », in *Knowledge modeling and expertise transfer*, Amsterdam, pp. 371-400.
- Dill S., Eiron N., Gibson D., Gruhl D., Guha R., Jhingran A., Kanungo T., Rajagopalan S., Tomkins A., Tomlin J. A. & Zien J. Y., (2003) "SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation", in *Proceedings of the 12<sup>th</sup> International World Wide Web Conference (WWW'03)*, ACM Press, Budapest, pp. 178-186.
- Domingue J. (1998) "Tadzebao and WebOnto : Discussing, Browsing, and Editing Ontologies on the Web", in *Proceedings of the 11th Knowledge Acquisition Workshop (KAW'98)*, Banff, Canada.
- Dwyer, F.R., Schurr P.H. et Oh S. (1987) "Developing Buyer-Seller Relationships," *Journal of Marketing* 51:(April) 11-27.
- Eisenhardt, K. & Tabrizi, B. (1995). "Accelerating adaptive processes: product innovation in the global computer Industry." *Administrative Science Quarterly*, Vol.40 (March), pp. 84-110.
- Eisenhardt K. (1989), "Building theories from case study research". *Academy of Management Review*. 14 (4), 532-550.
- Enjalbert P. (2005), « L'extraction d'information » in *Sémantique et traitement automatique du langage naturel*, Enjalbert P. (Ed.), *Traité IC2, Cognition et Traitement de l'Information*, Hermès, Paris, pp. 309-334.
- Ethiraj S. & Levinthal D. (2002), "Modularity and Innovation in Complex systems." Wharton Web Site.
- Favret-Saada J. (1977), *Les mots, la mort, les sorts*. Paris, Gallimard.

- Fernandez M., Gomez-Perez A; et Juristo N. (1997), “Methontology: from ontological art towards ontological engineering”, in *Proceedings of the Spring Symposium Series on Ontological Engineering (AAAI’97)*, AAAI Press, Stanford, CA.
- Fichman R. G. et Kemerer C. F. (1997), “The Assimilation of Software Process Innovations: An Organizational Learning Perspective”, *Management Science*, vol. 43(1), pp. 1345-1363.
- Fixson, K. S. (2003). *The Multiple Faces of Modularity - A Literature Analysis of Product Concept for Assembled Hardware Products*. Working Paper - University of Michigan / IMVP.
- Foray D. (2004), *Economics of Knowledge*. MIT Press. Cambridge MA.
- Ford D. (1980) “The Development of Buyer-Seller Relationships in Industrial Markets”, *European Journal of Marketing*, Volume 14 Issue: 5/6, 339 – 353.
- Ford D., Gadde L.-E., Hakansson H. et Snehota I. (2003), *Managing Business Relationships*. John Wiley & Sons Ltd. Chichester.
- Ford D., Gadde L.-E., Håkansson H., Lundgren A., Turnbull P., Wilson D., et Snehota I. (1998), *Managing Business Relationships*. John Wiley and Sons Ltd. Chichester.
- Forgues B. et Vandangeon-Derumez (2003), « Analyses longitudinales » in *Méthodes de Recherche en Management*, R. A. Thiétart, 2<sup>ème</sup> édition, Dunod, Paris.
- Fourcade F. (2004), *Les stratégies modulaires des équipementiers automobiles : définition, enjeux, méthodologies de conception et modèles financiers*, Thèse de Doctorat en Sciences de Gestion , École Polytechnique.
- Franke N. et von Hippel E. (2003), “Satisfying Heterogeneous User Needs via Innovation Toolkits: The Case of Apache Security Software.” *Research Policy*, 32, no. 7:1199–1215.
- Franke, N., et Shah S. (2003), “How Communities Support Innovative Activities: An Exploration of Assistance and Sharing Among End-Users.” *Research Policy* 32, no. 1: 157–178.
- Freeman C.(1982), *The Economics of Industrial Innovation, Second Edition*, Frances Printer, London.
- Fuchs C. (1993) *Linguistique et Traitements automatiques des langues*, Hachette, Paris.
- Garel G. (1994), « Réduction du temps de conception, concourance, et savoirs professionnels : les cas de l'emboutissage dans les projets automobiles », Thèse de Doctorat en Sciences de Gestion, Ecole Polytechnique, Paris.
- Garel G. (1998), Mémoire pour l’Habilitation à diriger des recherches, IAE, Université Paris 1, Panthéon Sorbonne, 3 juin.
- Garud R. et Kumaraswamy, A. (1995), “Technological and Organizational Designs for Realizing Economies of Substitution”, *Strategic Management Journal* 16 (Summer), 93-109.

- Gawer A. (2000), "The organization of platform leadership: an empirical investigation of intel's management process aimed at forstering complementary innovation by third parties", PhD Dissertation, Michigan Institute of Technology.
- Gawer A. & Cusumano M., (2002). *Platform leadership*. Harvard Business School Press, Cambridge, MA.
- Gawer A. et Henderson R. (2007), "Platform owner entry and innovation in complementary markets: Evidence from Intel," *Journal of Economics & Management Strategy*, 16, 1-34
- Giard V. et Midler C. (1993), (sous la direction de), *Pilotages de projet et Entreprises, Diversités et convergences*, ECOSIP, Economica. Paris.
- Gilmore J. H. & Pine B. J. (1993), *Mass Customization*. Harvard Business School Press. Boston, MA.
- Girin J. (1990), « L'analyse empirique des situations de gestion : éléments de théorie et de méthode ». In (s/d) Martinet, *Épistémologies et sciences de gestion*. Éditions Economica. Paris.
- Girin J. (1981), « Quel paradigme pour la recherche en gestion ? », *Economie et société*, série Sciences de gestion, n°2, décembre, pp 1872-1889.
- Girin J. (1979) *Karl L. Popper ou le conventionnalisme méthodologique*, texte non publié disponible à l'adresse URL : <http://crg.polytechnique.fr/incunables/Popper.html>.
- Girin J. (1983), *Les situations de gestion*, texte non publié disponible à l'adresse URL : <http://crg.polytechnique.fr/incunables/Situations.html>.
- Girin J. (1989), "L'opportunisme méthodique dans les recherches sur la gestion des organisations », Communication à la journée d'étude « La recherche-action en action et en question », *AFCET, Collège de systémique*, Ecole Centrale de Paris, mars.
- Glaser B. & Strauss, A. (1967), *The discovery of grounded theory: strategies for qualitative research*, Chicago, Aldine.
- Goffman E. (1974), "Field Research", *Journal of Contemporary Ethnography*, 18 (2), pp. 21-132.
- Grimaldi R. et Torrisi S. (2001), "Codified-Tacit and General-Specific Knowledge in the Division of Labor among Firms: a Study of the Software Industry", *Research Policy*, 30, 1425-1442.
- Grönroos C. (1990), *Service Management and Marketing : Managing the moments of Truth in Service Competition*. Lexington Books, New York.
- Grove A. (1996), *Only the Paranoid Survive: How to Exploit the Crisis Points that Challenge Every Company and Career*. Doubleday. New York
- Gruninger M. et Fox M. S. (1995), "Methodology for the design and evaluation of ontologies", in *Proceedings of the Workshop on Basic Ontological Issues on Knowledge Sharing (IJCAI'95)* Montréal.

- Guarino N. (1992), "Concepts, Attributes and Arbitrary Relations : Some linguistic and ontological criteria for structuring knowledge bases", in *Data and Knowledge Engineering*, 8(3), 1992, pp. 249-261.
- Hagg I. et Johanson J. (1983), *Firms in Network*, Business & Social Research Institute, Stockholm.
- Hakansson H. (ed) (1987), *Industrial Technological Development: A Network Approach* London Croom Helm.
- Hakansson H (ed) (1982), *International Marketing and Purchasing of Industrial Goods*, Chichester: Wiley.
- Hakansson H. (1989), *Corporate Technological Behaviour: Co-operation and networks*, London: Routledge.
- Hakansson H. et Snehota I. (1995) "Relationships in business". In Hakansson, H. & I. Snehota (Eds.), *Developing relationships in business networks* (pp. 1– 23). London: Routledge.
- Hakansson H. et Snehota I. (1989), "No business is an island". *Scandinavian Journal of Management*, 5(3), 187– 200.
- Hakansson H. et Waluszewski A. (eds) (2007), *Knowledge and Innovation in Business and Industry: The Importance of Using Others*. Routledge. London.
- Hakansson H. et Waluszewski A. (2002), *Managing Technological Development*. Routledge. London.
- Hallen L. et Sandström M. (1988), "Relationship Atmosphere in International Business," in Peter W. Tumbull and Stanley Paliwoda (eds.), *Research Developments in International Marketing*, Manchester: Manchester School of Management.
- Hand, J. (2001) *Evidence on the Winner-takes-all Business Model: The Profitability Returns-to-scale of Expenditures on Intangibles Made by U.S. Internet Firms, 1995-2001*, Working Paper, University of North Carolina at Chapel Hill.
- Handschuh, S., Staab S. et Ciravegna F. (2002), "S-CREAM - Semi-automatic CREAtion of Metadata", in *Proceedings of the 13th International Conference on Knowledge Engineering and Management (EKAW'02)*, LNCS 2473, Springer-Verlag, Madrid, Espagne, pp. 126-144.
- Harhoff, D. (1996), "Strategic Spillovers and Incentives for Research and Development." *Management Science* 42, no. 6: 907-925.
- Harhoff, D., J. Henkel, et von Hippel E. (2003), "Profiting from Voluntary Information Spillovers: How Users Benefit by Freely Revealing Their Innovations." *Research Policy* 32, no. 10: 1753-1769.
- Hatchuel, A. (1997), "Fondement des savoirs et légitimité des règles", in J. P. Dupuy et P. Livet *Les limites de la rationalité. Tome 2. La Découverte*. Paris

- Hatchuel, A. (1994), "Les savoirs de l'intervention en entreprise", *Entreprises et Histoire*, 7, pp. 59-75.
- Hatchuel, A., et Molet, H. (1986), "Rational Modelling in Understanding Human Decision Making." *European Journal of Operations Research*, 24, 178-186.
- Hatchuel, A., Weil, B., et Le Masson, P. (2002), « De la gestion des connaissances aux organisations orientées conception », *Revue Internationale des Sciences Sociales* 2002/1 - N° 171 p 29 à 42.
- Henderson, R. & Clark, K. (1990), "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the failure of Established Firms", *Administrative Science Quarterly*, 35, pp. 9-30.
- Hennessy J.L. et Patterson D.A. (1990), *Computer architecture: a quantitative approach*, Morgan Kaufmann Publishers Inc., San Francisco.
- Herstatt, C., et von Hippel E. (1992), "From Experience: Developing New Product Concepts via the Lead User Method." *Journal of Product Innovation Management* 9, no. 3: 213–222.
- Horn F. (2000), *L'économie du logiciel*, Thèse de doctorat d'Economie Industrielle, Université de Lille I.
- Horn F. (2004), *L'économie des logiciels*, La Découverte, coll. Repères, n° 381, Paris.
- Horn F. (1999), "Diversité des informations traitées par des moyens informatiques, standardisation optimale et acteurs du processus de standardisation", *Communications et stratégies* n°33, pp. 85-117.
- Hughes T. P. (1983), *Networks of Power: Electrification in Western Society, 1880-1930*. The Johns Hopkins University Press. Baltimore.
- Iansiti M. & MacCormack A. (1997), « Developing Products on Internet Time », *Harvard Business Review*, vol. 75, pp 108-117.
- Iansiti M. (1995). "Shooting The Rapids: Managing Product Development in Turbulent Environments", *California Management Review*, 38 (1, Fall), pp. 37-58.
- IBM (2006), Unstructured Information Management Architecture (UIMA), *SDK User's Guide and Reference*, 364 p., disponible à l'adresse URL : [http://dl.alphaworks.ibm.com/technologies/uima/UIMA\\_SDK\\_Users\\_Guide\\_Reference.pdf](http://dl.alphaworks.ibm.com/technologies/uima/UIMA_SDK_Users_Guide_Reference.pdf).
- Johanson J. et Mattsson L.-G. (1987), "Interorganizational Relations in Industrial Systems: A Network Approach Compared with the Transaction-Cost Approach", *International Studies of Management Organization*, Vol. XVII, No. 1, pp. 64-74.
- Johanson J. et Mattsson L.-G. (1994), "The Markets-as-Networks Tradition in Sweden." In: Laurent G., Lilien G. L. & Pras B. (eds.), *Research Traditions in Marketing*. Boston: Kluwer Academic Publishers, pp. 321-343.

- Johnston W. J. et Bonoma T.V. (1981), "The Buying Center: Structure and Interaction Patterns." *Journal of Marketing*, 45 (Summer), 143-156.
- Journé B. (2005), « Etudier le management de l'imprévu : méthode dynamique d'observation in situ. » *Finance, Contrôle, Stratégie* vol. 8 (4) 63-91.
- Katz R. et Allen T. J. (1988), « Organizational issues in the introduction of new technologies. » Ralph Katz, ed. *Managing Professionals in Innovative Organizations*, Ballinger, Cambridge, MA, 442–456.
- Koenig G. (1993), "Production de la connaissance et constitution des pratiques organisationnelles." *Revue de l'AGRH*, 9. pp. 63-78.
- Korson T. et Vaishnavi V. (1986), "An empirical study of the effects of modularity on program modifiability", *Proceedings of the first Workshop on Empirical Studies of Programmers*, 168- 186. Washington, D.C.
- Lakhani K. R. et von Hippel E. (2003), "How Open Source Software Works: 'Free' User-to-User Assistance." *Research Policy* 32, n° 6: 923-943.
- Lambe C.J., Spekman R.E. et Hunt S.D. (2000), "Alliance Competence, Resources, and Alliance Success: Conceptualization, Measurement, and initial Test" *Journal of the Academy of Marketing Science* vol 30 (2), 141-158.
- Langlois R.N. & Robertson P.L. (1992), "Networks and innovation in a modular system: Lessons from the microcomputer and stereo component industries", *Research Policy* 21 pp. 297-313.
- Langlois R.N. (2002), "Modularity in technology, Organization, and Society", *Journal of economic Behavior and organization*. Volume 49, Issue 1, September 2002, pp. 19-37.
- Latour B. (1994), « Pasteur et Pouchet: Hétérogénéité de l'Histoire des Sciences », in Serres, M. (ed.) *Éléments d'Histoire des Sciences*, Bordas, Paris, pp 423-445.
- Laublet P. (2007), « Web Sémantique et Ontologies », in *Nouvelles technologies cognitives et concepts des sciences humaines et sociales*, Volume 1, Humanités Numériques, Hermès, Paris.
- Law J. et Callon M. (1992), "The Life and Death of an Aircraft: a Network Analysis of Technological Change", pp. 29-52, in Bijker et Law (eds.) *Shaping Technology/ Building Society. Studies in Sociotechnical Change*, MIT Press, Cambridge MA.
- Le Masson P., Weil B. et Hatchuel A. (2006), *Les processus d'innovation - Conception innovante et croissance des entreprises*, Hermès. Paris.
- Lenfle S. (2001), « Compétition par l'innovation et organisation de la conception dans les industries Amont. Le Cas d'Usinor ». Thèse de Doctorat en sciences de gestion, Université de Marne-la-Vallée.
- Levinthal D. A. et March J.G (1993), "The myopia of learning" *Strategic Management Journal*, vol. 14, 2, p. 95-112.

- Levy M., et Jouillet J.-P. (2006), *L'économie de l'immatériel, la croissance de demain*, Rapport de la Commission sur l'Economie de l'Immatériel, Ministère de l'économie, des finances et de l'industrie Paris.
- Lewin K. (1951), *Field Theory in Social Science*, Harper and Row, New York.
- Lilien G. L., Morrison P. D., Searls K., Sonnack M., et von Hippel E. (2002), "Performance Assessment of the Lead User Idea-Generation Process for New Product Development." *Management Science*, 48, no. 8: 1042–1059.
- Llory M. (1996), *Accidents industriels le coût du silence. Opérateurs privés de parole et cadres introuvables*. L'Harmattan, Paris.
- Lundin R.A. & Wirdenius H. (1990), "Interactive research", *Scandinavian Journal of Management*, vol. 6 n°2, pp. 125-152.
- Lüthje, C., Herstatt, C. et von Hippel, E. (2002), « The Dominant Role of Local Information in User Innovation: The Case of Mountain Biking ». *Working Paper*, MIT Sloan School of Management.
- MacCormack A., Verganti R. & Iansiti M. (2001), "Developing Products on "Internet Time" : The Anatomy of a Flexible Development Process", *Management Science*, vol. 47, n°1, pp. 133-150.
- Mansfield E. (1972), "Contribution of R&D to economic growth in the United States." *Science* 175 (4021), 477-486.
- Martinet A.C. (1990), *Epistémologies et sciences de gestion*, Paris, Economica.
- Matthews J. (1985), *Public Access to Online Catalogs* 2<sup>nd</sup> Edition, New York: Neal-Schuman.
- Mauboussin M. J., Schay A, et Kawaja S. G. (2000), "Network to Net Worth, The Rising Role of the Replacement Cycle", Credit Suisse First Boston, Equity Research, *Frontiers of Finance*, vol. 5 (May 11, 2000), disponible à l'adresse URL : <http://www.capatcolumbia.com/Articles/FoStrategy/Ni1924.pdf>.
- Mauss M. (1923-1924), « Essai sur le don. Forme et raison de l'échange dans les sociétés archaïques », *L'Année Sociologique*, seconde série, 1923-1924. pp. 53-72.
- Mead C. et Conway L. (1980), *Introduction to VLSI Systems*, Addison-Wesley. Reading, MA.
- Menard S.(1991). *Longitudinal research*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-076. Newbury Park, CA: Sage.
- Michel D., Salle R. et Valla J.P., Eds. (1996), *Marketing Industriel*, Economica, Paris.
- Midler C., (1993), *L'auto qui n'existait pas, management des projets et transformation de l'entreprise*, InterEditions, Paris.

- Midler C. & al. (1996), "Du management de projet aux nouvelles rationalisations de la conception", *Cahiers du CRG*, n°13, Septembre.
- Mikkola J. H. (2003), "Modularity, component outsourcing, and inter-firm learning." *R&D Management*, 33(4): 439-454.
- Mikkola J. H. (2006), "Capturing the Degree of Modularity Embedded in Product Architectures" *Journal of Product Innovation Management*, 23(2): 128-146.
- Minel J.-L. (2003), *Filtrage sémantique. Du résumé à la fouille de textes*, Hermès, Paris.
- Miles et Huberman (1984), *Qualitative Data Analysis: A Sourcebook of New Methods*, Thousand Oaks, Sage Publication.
- Miles et Huberman (1994), *Qualitative Data Analysis*, Thousand Oaks, Sage Publication.
- Millier P. (2000), "Intuition can help in segmenting industrial markets", *Industrial Marketing Management*, 29, 147–155.
- Moisdon J. (1984), "Recherche en Gestion et intervention." *Revue Française de Gestion*. n° 47-48, pp. 61-73.
- Möller K. (1986), "Buying behaviour of industrial components: inductive approach for descriptive model building", *Research in International Marketing*, P. Turnbull, S.Paliwoda (Eds.), Croom Helm, 79-132.
- Morrison P. D. (1994), "A study of the relationship of leading edge users to opinion leaders", *Workingpaper n°94-001*, February, University of New South, Australie.
- Morrison P. D. (1995), *A Framework for Studying the Adoption of Technological Innovations by Organizations and the Role of Leading Edge Users in the Process*, Unpublished Doctoral Dissertation; Australian Graduate School of Management, UNSW.
- Morrison P. D., Midgley D. F. et Roberts J.H. (1994), "The Role of Leading Edge Users in the Adoption of Technological Innovations by Organizations," *Asia Pacific Advances in Consumer Research*, Vol. 1.
- Morrison P. D., Roberts J.H. et Midgley D.F. (2004), "The Nature of Lead Users and Measurement of Leading Edge Status." *Research Policy*, 33, n° 2: 351–362.
- Morrison P. D., Roberts J. H. et von Hippel E. (2000), "Determinants of User Innovation and Innovation Sharing in a Local Market." *Management Science*, 46, n° 12: 1513–1527.
- Naude P. et Turnbull P. W. (1998), *Network Dynamics in International Marketing*. Oxford: Elsevier Science Ltd.
- Nelson R. (1982), "The Role of Knowledge in R&D Efficiency," *The Quarterly Journal of Economics*, MIT Press, vol. 97(3), pp. 453-70.

- Nelson R. et Winter S. (1977), "In search of a useful theory of innovation", *Research Policy*, vol. 6, Issue 1, 36-76.
- Nobeoka K. et Cusumano M. (1995), "Multiproject strategy, design transfer, and project performance: a survey of automobile development projects in the US and Japan" in: *Engineering Management, IEEE Transactions* Nov. Vol. 42, Issue: 4, 397-409.
- Noe, T., Parker, G. (2000), « Winner take all: Competition, strategy, and the structure of returns in the internet economy », *Working Paper*, Tulane University.
- Nonaka I., (1994), "A Dynamic Theory of Organizational Knowledge Creation", *Organization Science*, Vol. 5, n°1, pp. 4-37.
- Nonaka, I. & Takeuchi, H. (1995). *The Knowledge-creating company*, Oxford, Oxford University Press.
- OCDE (1998), *Le secteur du logiciel : un profil statistique pour certains pays de l'OCDE*, OCDE, Paris.
- OCDE (2002), *Perspectives des technologies de l'information de l'OCDE: Les TIC et l'économie de l'information Edition 2002*, OECD Publishing, Paris.
- Orlikowski, W.J. (2005), "Material Works: Exploring the Situated Entanglement of Technological Performativity and Human Agency" *Scandinavian Journal of information Systems*, 17(1):183–186.
- Orlikowski, W.J. et Baroudi, J.J. (1991), "Studying Information Technology in Organizations: Research Approaches and Assumptions," *Information Systems Research*, (2:1), March, 1-28.
- Paris T. (2002), *Le droit d'auteur : l'idéologie et le système*, Lavoisier, Paris.
- Parnas D. (1972), "On the criteria to be used in decomposing systems into modules". *Communications of the ACM*, 15 :1053–1058.
- Parnas D., Clements P. et Weiss D. (1985), "The Modular Structure of Complex Systems", *IEEE Transactions on Software Engineering*, March Volume: SE-11, Issue: 3 pp. 259- 266.
- Pavitt K. (2002), "System integrators as "post-industrial" firms", *DRUID Conference*, Copenhagen Business School.
- Pavitt K. (1987), "The Objectives of Technology Policy." *Science and Public Policy*, 14, n°4 (August): 182-88.
- Pine J. B. II. (1999), *Mass Customization*, Harvard Business School Press, Boston.
- Popov B., Kiryakov A., Manov D., Kirilov A., Ognyanoff D. & Goranov M. (2003), "Towards Semantic Web Information Extraction", in *Proceedings of the Human Language Technologies Workshop (ISWC'03)*, Sanibel, Floride, pp. 1-22.
- Popper K. (1973), *Logique de la découverte scientifique*, Payot. Paris.

- 
- Prié Y. & Garlatti S., (2004), « Méta-données et annotations dans le Web sémantique », in *Le Web sémantique*, Charlet J., Laublet P. & Reynaud C. (Eds), Hors série de la *Revue Information -Interaction - Intelligence* (I3), 4(1), Cépaduès, Toulouse, pp. 45-68.
- Raulet-Croset N. (2003), « La méthode ethnographique appliquée aux ressources humaines », in Allouche J. (ed.) *Encyclopédie des ressources humaines*, Paris, Vuibert, pp. 878-887.
- Raymond E. (1999), *The Cathedral and the Bazaar* (revised edition), O'Reilly, 2001. Sebastopol, CA.
- Rice R. E. et Rogers, E.M. (1980), "Reinvention in the innovation process". *Knowledge* 1(4) 499–514
- Rogers E. M. (1995), *Diffusion of Innovation*, fourth edition. Free Press. New-York.
- Rogers E.M. (1983), *Diffusion of Innovations*. Free Press. New York.
- Rogers E. M. (1962), *Diffusion of Innovations*, Free Press, Glencoe.
- Rosenberg N. (1982), *Inside the black box: Technology and Economics*. Cambridge University Press, Cambridge, MA.
- Royce W.W. (1970), "Managing the development of large software systems: Concepts and techniques", *Procedures WESCON*, Western Electric Show and Convention, Los Angeles. Reprinted 1989 in *Proc. 11<sup>th</sup> Int. Conf. of Software Engrng.* Pittsburgh USA.
- Salle R. et Sylvestre H. (1992), *Vendre à l'industrie*, Editions Liaisons, Paris.
- Sanchez R. (1995), « Strategic flexibility in product competition », *Strategic Management Journal*, Summer Special Issue, 16, 135-159.
- Sanchez R. (1994), « Towards a science of strategic product design », paper presented at the Second International Product Development Conference on New Approaches to development and Engineering, 30-31 May, 1994, Gothenburg, Sweden.
- Sanchez R. & Mahoney J. (1996), "Modularity, flexibility, and Knowledge Management in product and organization design", *Strategic Management Journal*, 17, pp. 63-76 (Winter special issue).
- Sanchez R. et Sudharashan D. (1993), « Real-time market research: Learning-by-doing in the development of new products », *Marketing Intelligence and Planning*, 11 (7), 29-38.
- Sanderson S. W. et Uzumeri V. (1990), *Strategies for New Product Development and Renewal: Design-based Incrementalism*, Rensselaer Polytechnic Institute, Center for Science and Technology Policy, Working Paper, May.
- Schein E.H. (1987), *The clinical perspective in fieldwork*. Thousand Oaks, CA: Sage.
- Schilling M. A. (2000), "Towards a general modular systems theory and its application to interfirm product modularity". *Academy of Management Review*, 25(2): 312-334.

- Schilling M.A. (2003), "Modularity in Multiple Disciplines - Commentary on Towards a General Modular Systems Theory", in Garud, R., Kumaraswamy, A. & Langlois, R.N. (eds.) *Managing in the Modular Age: Architectures, Networks and Organizations*, pp. 203-214.
- Schilling M.A. et Steensma, H.K. (2001), "The use of modular organizational forms: an industry-level analysis", *Academy of Management Journal*, Vol. 44, n°6, 1149-1168.
- Schmookler J. (1966), *Invention and Economic Growth*. Cambridge: Harvard University press.
- Schultz D. J., (1997), *IEEE Standard for Developing Software Life Cycle Processes*, Technical Report 1074-1997, IEEE Computer Society, New-York.
- Schumpeter J. (1947), *Capitalisme, socialisme et démocratie*, Paris, Bibliothèque historique.
- Segelod E. et Jordan G. (2004), "The use and importance of external sources of knowledge in the software development process", *R & D Management* Num. 34, 3, pp. 64-83.
- Shapiro C. & Varian H. R. (1999), "The art of standards war", *California management review*, 41 (2): pp. 8-32.
- Shaw B. (1985), "The role of the interaction between the user and the manufacturer in medical equipment innovation." *R&D Management*, 15 (4), 283-292.
- Simon E. (1996), "Innovation and Intellectual Property Protection: The Software Industry Perspective." *Columbia Journal of World Business* 31, n°1: 30-37.
- Simon H. (1962), *The architecture of complexity*, Cambridge, MIT Press.
- Simon H. (1991), *Sciences des systèmes, sciences de l'artificiel*, Dunod, Paris.
- Sowa J. (2000), *Knowledge Representation: Logical, Philosophical and computational foundations*, Brooks Cole Publishing Co., Pacific Grove.
- Starr K. (1965), "Modular Production - A new Concept", *Harvard Business Review*, Nov. Dec., pp. 131-141.
- Steinmueller W.E. (2002), "Collaborative innovation: rationale, indicators and significance." In: J. de la Mothe and A.N. Link (eds), *Networks, Alliances and Partnerships in the Innovation Process*; Boston: Kluwer, pp. 29-43.
- Stryszowski P. (2008), "Innovation in the Software Sector", *Seminar on innovation and competitiveness policies*. Conférence OCDE, Genève, 14/02.
- Tedd L.A. (1994), "OPACs through the Ages" *Library Review* 43, No. 4, 27-37.
- Teece D. (1992), "Competition, cooperation and innovation", *Journal of Economic Behavior and Organization*, 18, pp. 1-25.
- Thiéart R.-A. et coll., (2003), *Méthodes de recherche en management*, Dunod, Paris.

- Thomke S. & von Hippel E., (2002), « Customers as Innovators , a New Way to Create Value », *Harvard Business Review*, April, pp. 74-81.
- Thomke S. H. (1998), “Managing Experimentation in the Design of New Products.” *Management Science*, 44, no. 6: 743–762.
- Thomke S. H., von Hippel E. et Franke R. (1998), “Modes of Experimentation: An Innovation Process and Competitive Variable.” *Research Policy*, 27, no. 3: 315–332.
- Truffle Venture (2005), *Truffle 100 2005*, accessible à l’URL : <http://www.truffle100.com/france/2005/chiffres.php>.
- Tyre M. et von Hippel E. (1997), “Locating Adaptive Learning: The Situated Nature of Adaptive Learning in Organizations.” *Organization Science*, 8, n°1, 71–83.
- Ulrich K. (1995), “The role of product architecture in the manufacturing firm”, *Research Policy*, 24, pp. 419-440.
- Ulrich K. & Eppinger S. D. (1995), *Product Design and Development*, New York, MacGraw-Hill.
- Upton D.M. (1997), “Process range in manufacturing: an empirical study of flexibility”, *Management Science*, vol. 43, Issue 8, 1079 - 1092.
- Urban G. L. et von Hippel E. (1988), “Lead User Analyses for the Development of New Industrial Products.” *Management Science* 34, no. 5: 569–82.
- Ushold M. et King M. (1995), “Towards a Methodology for Building Ontologies”, in *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing*, held in conjunction with IJCAI-95, Montreal, August 20-25.
- Utterback J. M. et Abernathy W.J. (1975), “A Dynamic Model of Process and Product Innovation.” *Omega* 3, n°6: 639–656.
- Valla J.-P. (1987), *L’approche interactive. Les travaux du groupe IMP en marketing industriel*, Working Paper IRE, Lyon, Octobre.
- Valla J.-P. (1986), “The French approach to Europe”. In: Turnbull P, Valla J-P, (eds) *Strategies for international industrial marketing*. London: Croom Helm, pp. 11 – 78.
- Van de Ven A. H. (1986), “Central problems in the management of innovation” *Management Science* Vol. 32, Issue 5 (May), pp. 590 – 607.
- Van de Ven A. H. et Huber G. P. (1990), “Longitudinal Field Research Methods for Studying Processes of Organizational Change” *Organization Science*, Vol. 1, n° 3, Special Issue: Longitudinal Field Research Methods for Studying Processes of Organizational Change, pp. 213-219.
- Van de Ven A. H. et Poole M. S. (1989), "Methods for studying innovation processes" in Van de Ven A., Angle H. & Poole M. S. *Research on the management of innovation*, New York, Harper Row, pp. 31-54.

---

Van de Ven, A. H., Polley, D., Garud, R., et Ventakaraman, S. (1999), *The Innovation Journey*, Oxford University Press, Oxford, UK.

Volberda H. W. (1996), «Toward the Flexible Form: How to Remain Vital in Hypercompetitive Environments », *Organization Science*, Vol. 7, n° 4. (Jul. - Aug.), pp. 359-374.

von Hippel E., (1990), «Task partitioning: An innovation process variable», *Research Policy* 19 (5), 407-418.

von Hippel E. (1986), «Lead Users: A Source of Novel Product Concepts», *Management Science*, vol. 32, n° 7: pp. 791–805.

von Hippel, E. (1988), *The sources of Innovation*, New York, Oxford University Press.

von Hippel E. (1994) «Sticky Information and the Locus of Problem Solving : Implications for Innovation», *Management Science*, vol. 40, 4, pp. 429-439.

von Hippel E. (2001), «Perspective: User Toolkits for Innovation», *Journal of Product Innovation Management*, vol.18, pp. 247–257.

von Hippel E. (1976), «The Dominant Role of Users in the Scientific Instrument Innovation Process», *Research Policy*, 5, n° 3: 212–39.

von Hippel E. (1977), «Transferring Process Equipment Innovations from User-Innovators to Equipment Manufacturing Firms», *R&D Management* 8, n° 1:13–22.

von Hippel E. (2005), *Democratizing Innovation*. The MIT Press, Cambridge MA.

von Hippel E. (1998), «Economics of Product Development by Users: The Impact of Sticky Local Information.» *Management Science* 44, no. 5: 629–644.

von Hippel E. (2001), «Perspective: User Toolkits for Innovation.» *Journal of Product Innovation Management*, 18: 247–257.

von Hippel E. (2002), « Vos clients conçoivent-ils vos nouveaux produits ? » *Entretien Cefrio*, disponible à l'adresse URL : [http://cefrio.qc.ca/pdf/Perspectives\\_article3.pdf](http://cefrio.qc.ca/pdf/Perspectives_article3.pdf).

von Hippel E. et Katz R. (2002), «Shifting Innovation to Users via Toolkits.» *Management Science*, vol. 48, n° 7: pp. 821–833.

von Hippel E. et von Krogh G. (2003), «Open Source Software and the “Private- Collective” Innovation Model: Issues for Organization Science.» *Organization Science*, 14, n°2: 209–223.

von Hippel E., Thomke, S. H. et Sonnack M. (1999), «Creating Breakthroughs at 3M.» *Harvard Business Review*, 77, n°5: 47–57.

von Krogh G. & von Hippel E. (2003), « Special Issue on Open Source Software development», Editorial in *Research Policy* 32/7 : pp. 1146 - 1157.

Voss C. (1985), "The role of users in the development of applications software", *Journal of Product Innovation Management*, vol. 2, 113-121.

Wacheux F. (1996), *Méthodes qualitatives et recherche en gestion*, Economica, Paris.

Weil B. (1999), *Conception collective, coordinations et savoirs, les rationalisations de la conception automobile*, Thèse de doctorat en Sciences de Gestion, Ecole Nationale Supérieure des Mines de Paris.

Woodall P. (2000), "Survey: The new Economy: Falling through the Net ?", *The Economist*, Sept. 23, 356, 8189.

Woolsey J. P. (1974), "777", *Air Transport World*, 22-31.

Yin R. (1994), *Case study research: Design and methods* (2<sup>nd</sup> ed.). Beverly Hills, CA: Sage Publishing.

Yin R. (2003), *Case Study Research. Design and Methods*. (3<sup>rd</sup> ed.) Beverly Hills, CA: Sage Publishing.

---

**Annexe :**  
**Web sémantique et notions**  
**liées.**

---

## SOMMAIRE DES ANNEXES

<b>Annexe</b>	<b>352</b>
<i>Sommaire des Annexes</i>	353
<i>Annexes : Web sémantique et notions liées.</i>	354
1. <i>L'annotation sémantique, définitions :</i>	354
2. <i>Le problème à résoudre : l'automatisation de la création des annotations en utilisant un moteur d'extraction d'information.</i>	357
2.1. <i>Création manuelle des patrons : le fonctionnement des moteurs d'extraction d'information (EI).</i>	357
2.2. <i>Apprentissage supervisé des patrons :</i>	358
3. <i>Le stockage des annotations et les bases de connaissance :</i>	358
4. <i>L'utilisation des annotations par les agents logiciels :</i>	359
5. <i>L'annotation et le Web Sémantique :</i>	359
6. <i>Les Ressources Terminologiques ou Ontologiques (RTO) : définitions.</i>	359
6.1. <i>Les Taxonomies :</i>	360
6.2. <i>Les Thesaurus :</i>	360
6.3. <i>Les Ontologies</i>	361
6.4. <i>Les RTO et l'annotation sémantique :</i>	362

---

## ANNEXES : WEB SEMANTIQUE ET NOTIONS LIEES.

(D'APRES AMARDEILH, 2007):

### 1. L'ANNOTATION SEMANTIQUE, DEFINITIONS :

Le terme annotation est défini comme une note critique ou explicative qui accompagne un texte – une note de lecture qu'on inscrit sur un livre. Les annotations seules ne signifient rien, elles sont considérées comme des métadonnées. Prié & Garlatti (2004) définissent une métadonnée comme une description normalisée attachée à une ressource identifiée (sur le Web notamment) et une annotation comme un commentaire libre situé à l'intérieur de la **ressource documentaire** (l'ensemble d'un document ou un fragment de celui-ci qui peut contenir du texte, de l'image, du son, de la vidéo).

Dans le cadre du Web Sémantique, une annotation descriptive, notamment lorsqu'elle s'intéresse à la structure logique du contenu d'un document, est le plus souvent appelée annotation sémantique. Le terme « sémantique » indique la volonté de faire émerger le sens d'un contenu et ce, de manière plus ou moins formelle selon les préceptes de la logique. Les annotations sémantiques ont donc pour objectif d'exprimer la « sémantique » du contenu d'une ressource afin d'en améliorer la compréhension, la recherche et donc la réutilisation par les utilisateurs finaux.

L'annotation sémantique consiste à ajouter (semi-)automatiquement des métadonnées structurées aux ressources documentaires du Web mais aussi des intranets des entreprises. Les annotations décrivent aussi bien le document dans son ensemble, comme son titre, son auteur, etc., que son contenu par des descripteurs provenant de vocabulaires contrôlés comme les thésaurus ou par des instances d'une base de connaissance. Ces annotations sont alors exploitables par les utilisateurs finaux d'une application donnée pour rechercher, partager, accéder, publier des documents, des métadonnées ou même de la connaissance.

Traditionnellement, **l'analyse linguistique d'un texte est découpée et organisée en différents niveaux** relevant de la structure interne des mots (morphologie), de leur organisation en groupes de mots et en phrases (syntaxe), de l'analyse du sens des mots et des phrases (sémantique) (Sowa 2000). Chaque niveau de l'analyse linguistique génère un ensemble

d'annotations (au sens général du terme, i.e. où de l'information est ajoutée au document textuel afin d'en caractériser le contenu). Ces annotations sont aussi appelées étiquettes dans la mesure où chaque niveau d'analyse attache directement un ensemble d'étiquettes prédéterminées à chaque unité textuelle traitée. Un exemple, issu du projet PressPro, d'annotations générées à chaque niveau d'analyse linguistique est donné dans la figure 27.

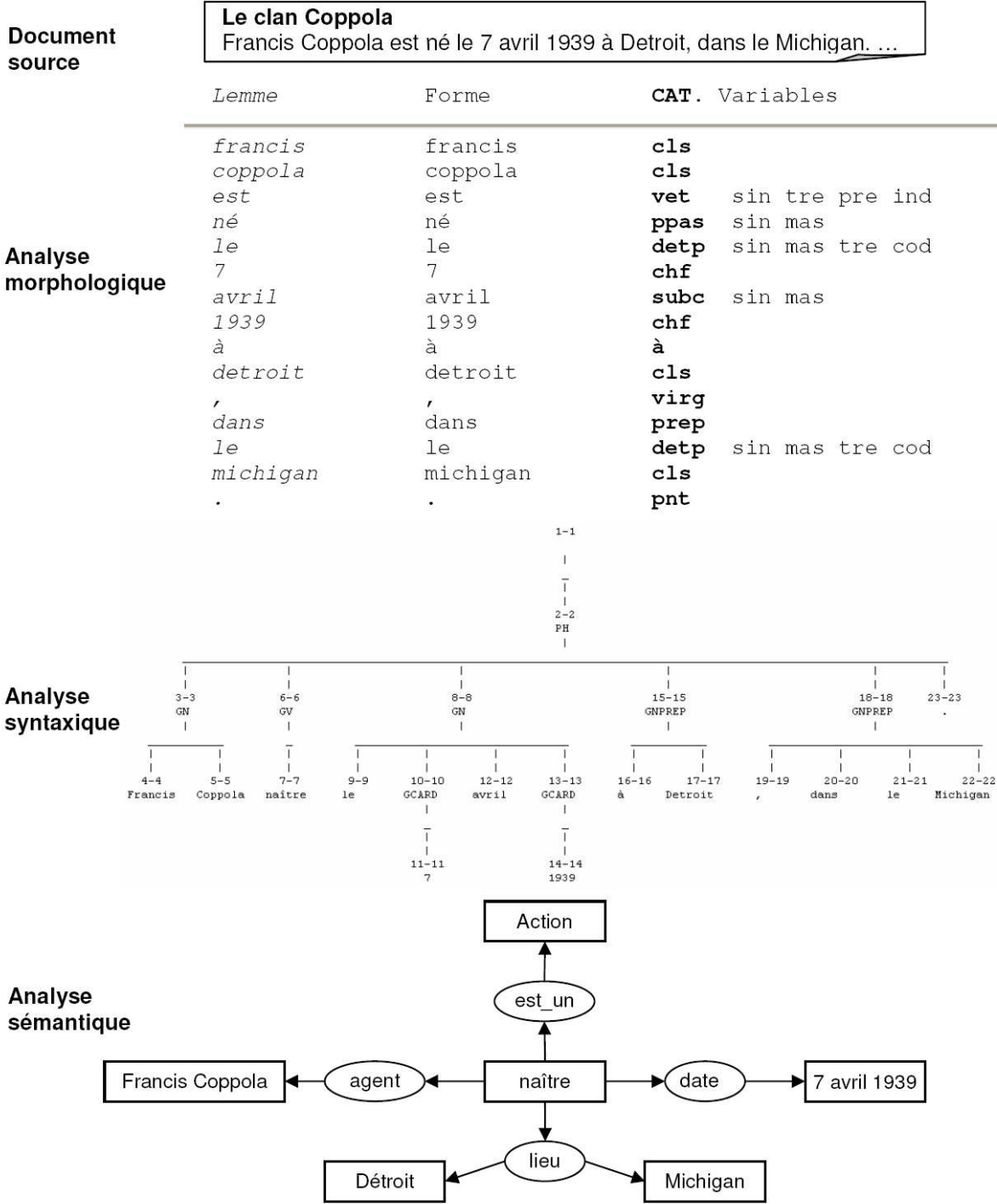


Figure 27 : Exemple d'annotations générées aux différents niveaux morphologique, syntaxique et sémantique d'une analyse linguistique (Amardeilh, 2007)

L'analyse sémantique est au cœur de tous les mécanismes de compréhension de la langue, permettant d'analyser, de traduire et d'interpréter les phrases et plus globalement les textes. Le lexique joue un rôle central dans la résolution des ambiguïtés et des exceptions liées à cette analyse sémantique – deux problèmes majeurs pour le TAL (Traitement Automatique du Langage). Par ailleurs, comparativement à la morphologie et la syntaxe, les travaux en sémantique n'ont pas atteint le même niveau de développement, car le niveau sémantique est beaucoup plus complexe à décrire et à formaliser. Aussi les réalisations opérationnelles sont plus rares, et concernent des applications très limitées où l'analyse sémantique se réduit à l'analyse d'un domaine parfaitement circonscrit (Fuchs, 1993).

Il faut disposer d'un format de représentation standard, consensuel et formel pour que les annotations puissent être créées, exploitées et maintenues par différents utilisateurs, qu'ils soient humains ou logiciels (Amardeilh, 2007). Une ontologie, telle que définie dans le cadre de l'Ingénierie des Connaissances, représente à la fois cet objet de consensus pour les humains et un objet formel permettant son exploitation par un agent logiciel (Laublet, 2007). Elle se compose de concepts (ou classes) et de propriétés (relations ou attributs) et de contraintes qui définissent et précisent l'utilisation de ces concepts et propriétés. Elle est décrite dans un langage formel de représentation des connaissances, tel que RDF(S) ou encore OWL. Une annotation dans ce contexte permet de relier le contenu du texte à des instances de concepts ou de propriétés décrits dans l'ontologie de référence, cf. Figure 28.

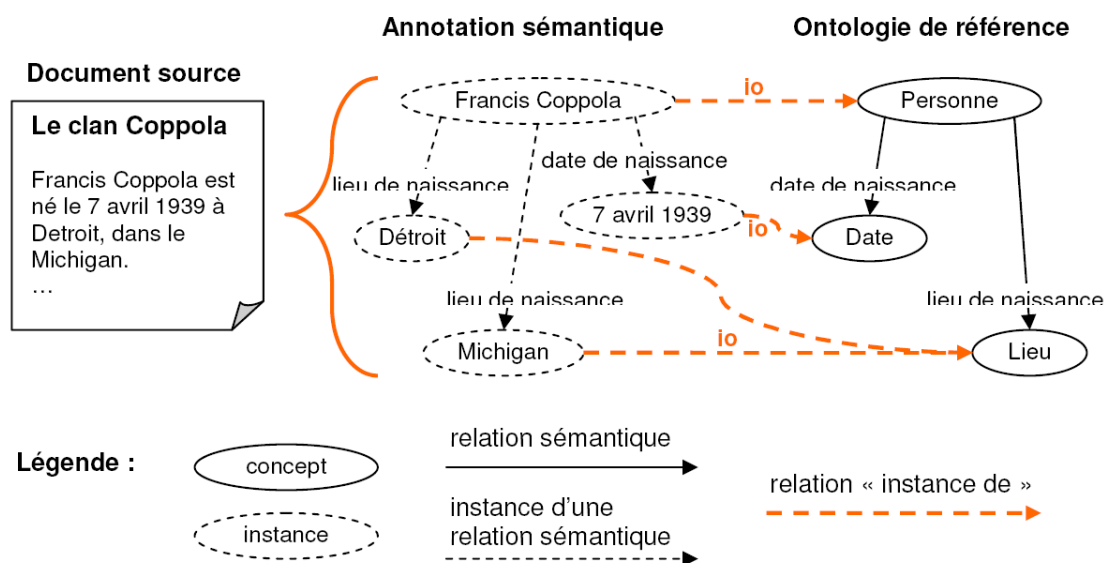


Figure 28 : Exemple d'une annotation sémantique orchestrée par une ontologie de référence (d'après Amardeilh, 2007).

---

Le fait d'utiliser la sémantique de ces concepts et relations telle que modélisée dans l'ontologie pour créer l'annotation, permet alors d'assigner cette sémantique au contenu de la ressource et donc de le rendre exploitable par des agents logiciels.

## **2. LE PROBLEME A RESOUDRE : L'AUTOMATISATION DE LA CREATION DES ANNOTATIONS EN UTILISANT UN MOTEUR D'EXTRACTION D'INFORMATION.**

Le traitement automatisé pour la création des annotations sémantiques est en réalité généralement semi-automatique. En effet, les traitements entièrement automatisés utilisent des algorithmes basés sur des modèles statistiques ou sur l'exploitation de la redondance dans un corpus de ressources (Dill et alii 2003). Mais leur perspicacité est limitée et leurs annotations souvent inexploitables par une application en entreprise. Par contre, les traitements semi-automatiques ont prouvé au cours des dernières années qu'ils pouvaient apporter une aide non négligeable à l'annotateur humain. Ces traitements semi-automatiques s'appuient généralement sur un moteur d'extraction d'information, dont la construction s'effectue selon deux méthodes : soit la création manuelle de ses patrons d'extraction (Popov et al. 2003), soit leur apprentissage supervisé (Hands Schuh, 2002) :

### **2.1. Création manuelle des patrons : le fonctionnement des moteurs d'extraction d'information (EI).**

Cette méthode fait fonctionner un moteur d'extraction d'information à partir d'un ensemble de règles définies par un utilisateur expérimenté, généralement un linguiste. Ces règles sont soit des expressions régulières (par exemple l'expression régulière d'une date sous la forme «jj/mm/aaaa») soit par des résultats d'analyses linguistiques plus poussées (par exemple si l'analyse syntaxique trouve un sujet représentant une entité nommée de type personne et un verbe signifiant la naissance et un complément de lieu représentant une entité nommée de type lieu, alors on peut annoter cette phrase comme étant le lieu de naissance de la personne du sujet). L'ensemble de ces règles est ensuite compilée dans un automate à états finis. L'automate parcourt l'ensemble du corpus documentaire et l'annote en fonction des règles qui sont déclenchées. Les annotations sont ensuite proposées à l'annotateur humain via une interface dédiée. Celui-ci doit alors les passer en revue pour les valider, les modifier ou les supprimer. L'analyse linguistique convient bien à l'annotation de textes semi ou non structurés. Par contre, la définition des règles

---

s'appuyant sur des expressions linguistiques particulières à un domaine donné, cette méthode est fortement dépendante de celui-ci.

## 2.2. Apprentissage supervisé des patrons :

Cette méthode permet de mettre en place un système capable d'apprendre à annoter un corpus donné. L'annotateur humain initie cet apprentissage en annotant manuellement un sous-ensemble du corpus. Ces annotations servent alors d'exemples au moteur d'extraction qui en déduit un ensemble de règles. Cette méthode est adaptée aux documents très structurés (issus des bases de données) ou semi-structurés (comme les pages Web) dans lesquels les constructions sont suffisamment simples et fixes pour que des expressions régulières linguistiques ou structurelles puissent être apprises facilement.

## 3. LE STOCKAGE DES ANNOTATIONS ET LES BASES DE CONNAISSANCE :

Les annotations peuvent être soit « embarquées » soit « débarquées » vis-à-vis de la ressource documentaire source. Une annotation est dite « embarquée » lorsqu'elle est ajoutée directement au contenu du document d'origine. A contrario, l'annotation est dite « débarquée » lorsqu'elle est stockée à l'extérieur du document source. Non seulement l'annotation elle-même doit être stockée, mais le lien avec la ressource annotée doit aussi être préservé. Généralement, ces annotations sont stockées sur des serveurs d'annotations qui peuvent être interrogés afin de retrouver les annotations d'une ressource donnée. L'avantage de l'annotation débarquée est qu'il devient possible d'annoter toute ressource documentaire, y compris celles dont l'application n'est pas propriétaire. Les annotations sémantiques peuvent être structurées et décrites formellement pour représenter de la connaissance sous la forme d'instances de concepts et de relations de l'ontologie de référence. Ces instances peuvent être stockées dans une base de connaissance, indépendamment de la ressource annotée, afin d'être réutilisées pour de nouvelles annotations d'autres ressources ou tout simplement pour être interrogées par les utilisateurs de cette base de connaissance. Dans ce cas, on parle de **peuplement d'ontologie** ou d'**enrichissement de base de connaissance**. Ces notions décrivent l'action d'ajouter de nouvelles instances à une base de connaissance contrainte par une ontologie, autrement dit à peupler une ontologie avec des instances de concepts et de relations. Les outils d'annotation sémantique peuvent donc peupler une ontologie existante à partir des annotations sémantiques créées pour une ressource donnée.

---

#### **4. L'UTILISATION DES ANNOTATIONS PAR LES AGENTS LOGICIELS :**

L'annotation sémantique trouve de nombreuses applications comme la recherche d'information sémantique, la catégorisation, la composition de documents, etc. (Prie et Garlatti, 2004). Enfin, plus le modèle de l'annotation est formalisé, plus les services proposés à partir de cette annotation peuvent devenir « intelligents ». En effet, les agents logiciels pourront inférer de la connaissance nouvelle, raisonner sur cette connaissance et ainsi améliorer les résultats de la recherche d'information ou bien dégager un sens implicite contenu dans le document d'origine (Laublet, 2007).

#### **5. L'ANNOTATION ET LE WEB SEMANTIQUE :**

L'objet du web sémantique est de décrire le contenu des ressources en les annotant avec des informations non ambiguës afin de favoriser l'exploitation de ces ressources par des agents logiciels (Prie et Garlatti, 2004). Or, les données actuelles du Web sont encore trop souvent écrites en langage naturel, car destinées aux humains. Le langage naturel étant par essence trop ambigu, des alternatives formelles et sémantiquement explicites doivent être mises en place pour lever les ambiguïtés du langage naturel. La tâche d'annotation pour le Web Sémantique consiste donc à prendre en entrée une ressource documentaire et fournir en sortie le même contenu enrichi par des annotations sémantiques basées sur des représentations de la connaissance plus ou moins formelles. Nous allons maintenant nous intéresser aux ressources terminologiques ou ontologiques (RTO) qui permettent de représenter la connaissance d'un domaine, puis nous verrons comment les annotations sémantiques relient un document source à ces RTO.

#### **6. LES RESSOURCES TERMINOLOGIQUES OU ONTOLOGIQUES (RTO) : DEFINITIONS.**

Bourigault et al. (2004) ont défini la notion de Ressources Terminologiques ou Ontologiques (RTO) à la croisée des domaines de la Terminologie et de l'Intelligence Artificielle, et plus particulièrement de l'ingénierie des connaissances. Cette notion regroupe plusieurs sortes de ressources, allant des index et glossaires jusqu'aux ontologies en passant par les bases de données lexicales et les thesaurus. Trois RTO principales permettent de représenter et de modéliser la connaissance d'un domaine : les taxonomies, les thesaurus et les ontologies.

## 6.1. Les Taxonomies :

L'objet d'une taxonomie est de conceptualiser les objets du monde et de les organiser hiérarchiquement les uns par rapport aux autres. Les sciences naturelles ont construit des taxonomies, au sujet de la botanique ou de la faune, comme celle présentée dans la Figure 29.

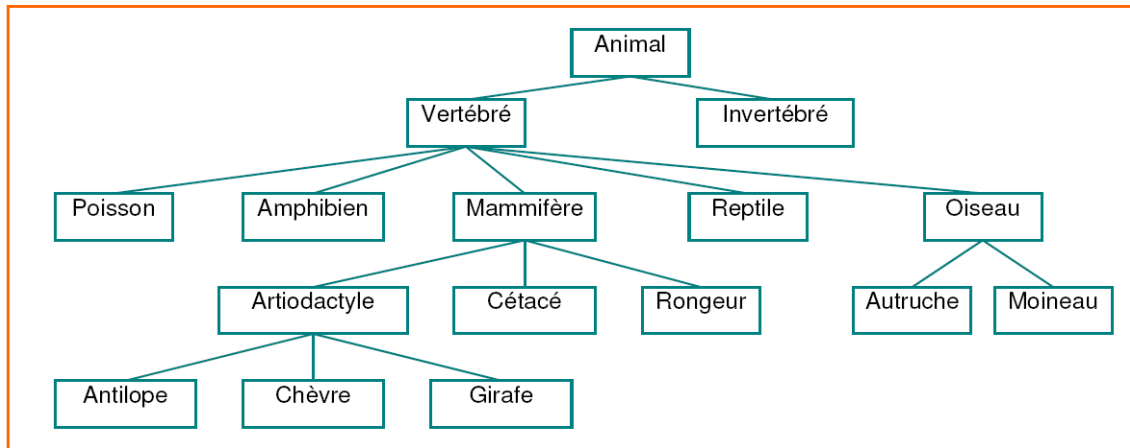


Figure 29 : Extrait d'une taxonomie sur la représentation simplifiée de la faune (Amardeilh, 2007)

## 6.2. Les Thesaurus :

Les domaines de l'Informatique documentaire et de la terminologie ont depuis longtemps mis au point des modélisations de ressources terminologiques comme les thesaurus. Un thesaurus est donc considéré comme un vocabulaire contrôlé et structuré dans lequel les relations entre les termes du domaine considéré sont clairement spécifiées formant ainsi un réseau terminologique (figure 30).

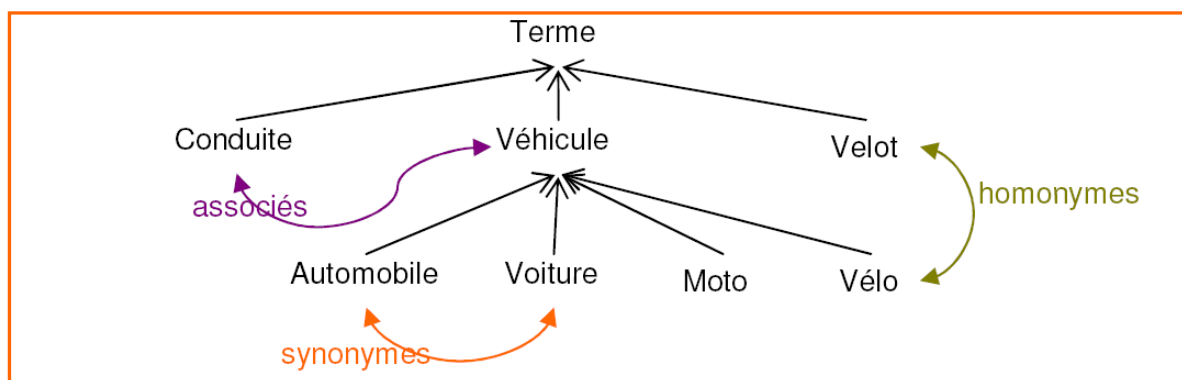


Figure 30 : Les différentes relations qui composent un thesaurus (Amardeilh, 2007).

---

L'objectif premier d'un thesaurus est de faciliter la recherche de document et de rendre l'indexation de documents consistante à l'aide de ses termes, appelés descripteurs.

### 6.3. Les Ontologies

Une ontologie donne une représentation de la connaissance propre à un domaine, elle fournit les moyens d'exprimer les concepts d'un domaine en les organisant hiérarchiquement et en définissant leurs propriétés sémantiques dans un langage de représentation des connaissances formel favorisant le partage d'une vue consensuelle sur ce domaine entre les applications informatiques qui en font usage. Définir des concepts et les relier entre eux par des relations sémantiques correspond au premier niveau d'une ontologie, le modèle conceptuel, inspiré des réseaux sémantiques, et plus encore des graphes conceptuels de Sowa (2000).

Les **concepts** (aussi appelées « classes ») représentent les objets, abstraits ou concrets, du monde réel. Ces concepts sont organisés en taxonomie. Dans la figure 31, le concept « Personnalité » est une sous-classe de « Personne ». Les **relations** représentent des interactions entre concepts permettant de construire des représentations complexes de la connaissance du domaine. Elles établissent des liens sémantiques binaires, organisables hiérarchiquement. Dans le domaine modélisé ci-dessous, les concepts « Personnalité » et « Film » sont reliés entre eux par la relation sémantique « réalise(Personnalité, Film) » dans laquelle « Personnalité » est le domaine et « Film » la portée (ou « range » en anglais). Les **attributs** correspondent à des caractéristiques, attachées à un concept et qui permettent de le définir de manière unique dans le domaine. Leurs valeurs sont littérales, i.e. de type primitif, comme une chaîne de caractère ou un nombre entier. Par exemple, un concept « Personne » peut avoir les attributs suivants : un « numéro de sécurité sociale », une « date de naissance », etc. Le terme propriété englobe les attributs et les relations, c'est-à-dire toutes les caractéristiques propres à un concept.

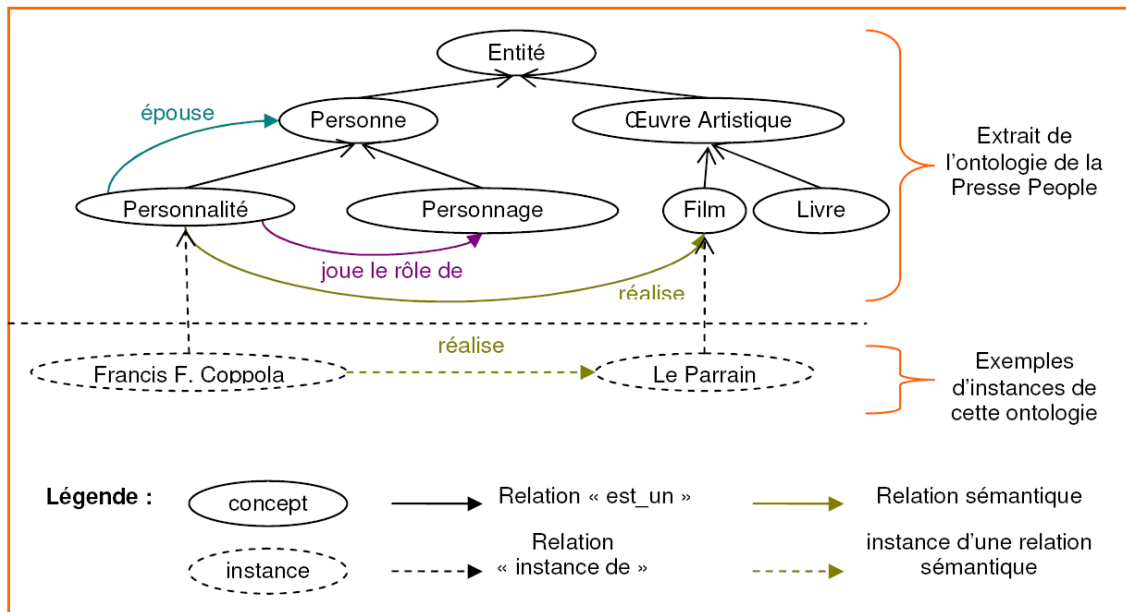


Figure 31 : Exemple d'une ontologie dans le domaine de la presse « People » (Amardeilh, 2007)

Les **instances** de concepts (aussi nommés individus) ne font pas à proprement parler partie de l'ontologie, mais plutôt de la base de connaissance. En effet, les bases de connaissances permettent de stocker les instances des concepts, mais aussi les instances de relations et les valeurs des propriétés en fonction des contraintes imposées par l'ontologie. Dans l'exemple de la Figure 31, « Francis F. Coppola » est une instance du concept « Personnalité » et une relation sémantique « réalise » est instanciée entre cette instance et celle du concept « Film », i.e. « Le Parrain ». La base de connaissance contiendra donc les informations *Personnalité(Francis F. Coppola)*, *Film(Le Parrain)* et *réalise(Francis F. Coppola, Le Parrain)*. L'action de définir et d'instancier une base de connaissance est appelée « peuplement d'ontologie ».

Par rapport aux taxonomies, et aux thesaurus, les ontologies (conceptualisations plus riches, plus complexes) autorisent la mise en place de mécanismes d'inférence pour la découverte et l'exploitation de la connaissance d'un domaine).

#### 6.4. Les RTO et l'annotation sémantique :

Bien que les thesaurus ne constituent pas une ontologie, ils peuvent permettre de constituer un premier niveau d'annotation sémantique. Dans ce cas, les annotations consistent en de simples pointeurs vers les termes du thesaurus, aussi appelés **descripteurs**. Elles fonctionnent alors

---

comme des entrées d'index améliorant les résultats des moteurs de recherche qui utilisent ce genre d'annotation.

Les ontologies de domaine, quant à elles, sont de plus en plus utilisées pour annoter des ressources documentaires. En fait, une même ressource peut être annotée par différentes ontologies de domaine en même temps, offrant ainsi différents points de vue sur un même contenu. Non seulement, les annotations sémantiques créées à partir d'une ontologie améliorent la recherche d'information, car elles peuvent contribuer à faire jouer des mécanismes d'inférence et de raisonnement, mais elles peuvent être combinées avec l'enrichissement d'une base de connaissance du domaine. En effet, la connaissance du domaine contenue dans le document peut être extraite à partir des annotations sémantiques afin d'être stockée et exploitée dans une base de connaissance contrainte par l'ontologie du domaine.

Les thésaurus et les ontologies offrent des accès aux contenus d'après des angles différents : celui du vocabulaire pour les thésaurus et celui de la conceptualisation d'un domaine pour l'ontologie. Ils ont donc des rôles complémentaires à jouer dans l'annotation sémantique comme nous allons le voir dans l'exemple suivant. Il s'agit d'un extrait de l'article « Le Clan Coppola » paru le 30/02/2003 dans le magazine ELLE, au sujet de la famille de Francis Ford Coppola (figure 32).

Annotons tout d'abord l'extrait avec un thésaurus géographique comme le TGN. Dans la figure 32, ces annotations correspondent aux mots surlignés en vert (foncé) :

- **Detroit** est une référence au descripteur ayant l'ID 7013547 dans le TGN et qui fait référence à la ville de Détroit dans l'état du Michigan.
- **Michigan** est une référence au descripteur ayant l'ID 7007520 et qui fait référence à l'état Michigan aux Etats-Unis
- **New York** est une référence au descripteur ayant l'ID 7007567 dans le TGN et qui fait référence à la ville de New York aux Etats-Unis.

## LE CLAN COPPOLA

Francis Coppola naît le 7 avril 1939 à Detroit, dans le Michigan. Il est le deuxième des trois enfants de Carmine et Italia Coppola. Son père, originaire de New York, est chef d'orchestre. Francis fera appel à lui pour composer la musique du « Parrain », en 1972. Sa mère, elle, est la fille du célèbre compositeur napolitain Francesco Pennino, auteur de l'opéra « Senza Mamma », dont un extrait figure dans « Le Parrain II ». Comédienne, elle joue dans plusieurs films de Vittorio De Sica, avant d'embrasser la carrière de « mamma ». Francis a de qui tenir ! [...]

Figure 32 : Extrait de l'article « Le Clan Coppola » paru dans le magazine ELLE, le 30/02/2003. (Amardeilh, 2007)

Puis, supposons qu'il existe une ontologie dans le domaine de la « Presse People » que nous pourrions utiliser pour annoter l'extrait ci-dessus. Cette ontologie contiendrait les concepts « Personnalité » et « Œuvre Artistique », des attributs comme « lieu de naissance », « date de naissance », « date de création », « profession », etc. et des relations telles que « est créé par(Œuvre Artistique, Personnalité) » ou « a lien parenté avec(Personnalité, Personnalité) ».

Les documents seraient alors annotés en fonction de ces éléments, à savoir :

- - Par les instances de concepts. Par exemple, une annotation sera créée avec « Francis Ford Coppola » qui représente une instance du concept « Personnalité » et avec « Le Parrain », instance du concept « Œuvre Artistique ».
- - Par les valeurs d'attributs. Par exemple, « 7 avril 1939 » peut être la valeur de l'attribut « date de naissance » qui sera attaché à l'instance « Francis Ford Coppola » de notre annotation.
- - Par les instances de relations. Par exemple, l'instance « Le Parrain » du concept Œuvre Artistique et l'instance « Francis Ford Coppola » du concept Personnalité peuvent être connectées par la relation « est-crée-par ». Le document sera annoté avec cette instance de relation « est créée par (Le Parrain, Francis Ford Coppola) ».