

## **Chapitre 4. Instrumentation de l'intertextualité et de l'intersubjectivité**

Dans le précédent chapitre, nous avons vu comment les méthodes en Sciences Humaines recommandaient la constitution de corpus comme « lieux » d'intersubjectivité et d'intertextualité. Nous allons maintenant nous interroger sur l'instrumentation de ces deux notions à l'aide de systèmes informatiques. Tout d'abord, nous verrons que ces notions sont au cœur des usages (dans les bibliothèques traditionnelles). Ensuite, nous étudierons comment elles ont été modélisées dans le cadre de systèmes hypermédia.

### **1. Au cœur des usages**

Instrumenter un travail n'est jamais une opération neutre. Comme le rappelle Bruno Bachimont [Bachimont99a] n'importe quel outil (du marteau au système informatique) détermine par sa structure des usages possibles (ce qui n'empêche pas des usages déviants). La question de l'adéquation de l'outil au travail à effectuer est donc déjà sensible au niveau d'une entreprise. Elle devient flagrante dans le cas de la recherche publique, pour laquelle, il serait tout à fait déplacé de la part d'une instance hiérarchique d'imposer l'utilisation d'un outil.

Aussi, certains pourraient être surpris par notre choix d'élaborer un modèle du travail à partir de textes méthodologiques (censés le guider) plutôt qu'à partir d'un travail « d'anthropologie » (immersion dans une communauté, observation et entretiens). Cependant, il faut constater que les études « anthropologiques » menées par d'autres équipes corroborent tout à fait l'utilisation de l'intertextualité et de l'intersubjectivité.

Ainsi, Kenton O'Hara et son équipe [OHaraEtAl98] ont étudié les activités documentaires effectuées par des doctorants durant une journée de travail. Ils décrivent le travail des usagers des bibliothèques universitaires comme allant bien au-delà de la traditionnelle recherche d'information bibliographique. En effet, devraient être considérées comme formant un tout :

- la lecture,
- l'annotation,
- la rédaction de fiches de lecture,
- la reproduction par photocopie,
- la consultation d'une bibliographie,
- la relecture,
- l'écriture de nouveaux documents (articles, mémoires, bibliographies...).

De même, Jacques Virbel et son équipe, à la suite d'une collaboration avec les grands lecteurs de la Bibliothèque Nationale de France, ont mis en évidence huit verbes d'action liés à la lecture active de documents [MazhoudEtAl95, Veron97] :

- Hiérarchiser (attribuer un niveau d'importance),
- Architecturer (typer des fragments),
- Contextualiser (borner le sens),
- Reformuler,
- Commenter (critique, idée associée),
- Documenter (ajouter un fragment pour en comprendre une autre),
- Corréler (relier deux fragments),
- Programmer (projeter une action : à traduire, à relire, à analyser...).

Enfin, dans le centre de documentation d'une entreprise, Andreas Paepcke [Paepcke96] a interviewé des ingénieurs sur leurs besoins et pratiques informationnelles. Il en ressort que même si la recherche d'information est centrale, celle-ci est corrélée avec trois actions nécessitant la communication entre des êtres humains, à savoir :

- la découverte,
- la gestion
- et le partage de ces informations.

Après de tels constats, on pourrait s'étonner du fait que très peu de systèmes de bibliothèques numériques permettent les interactions sociales [Tochtermann96] et l'enrichissement par le lecteur [GohEtLeggett00, RoscheisenEtAl95, NanardEtNanard00]. Cependant, comme nous allons le voir, ces thèmes ont été en partie traités par une autre communauté : celle des Hypermédia.

## 2. Panorama des systèmes hypermédia

Sans prétendre à l'exhaustivité, nous aimerions donner une vue aussi représentative que possible de différents types de modèles hypermédia. La documentation des modèles sous-jacents étant souvent insuffisante, nous avons dû effectuer une « rétro-conception » à partir des bribes de modèles donnés par les auteurs, à partir d'exemples d'utilisation, ainsi que de copies d'écran.

### a. Xanadu®

*Xanadu*® est à juste titre appelé « the original hypertext project ». En effet, il s'agit du projet que poursuit, depuis les années 1960, Ted Nelson (alors licencié en philosophie et maître en Sociologie), inventeur des termes « hypertexte » et « hypermédia ». L'idée originale consiste à lire deux textes en parallèle en exhibant les relations d'analogie (indiquées par le lecteur) ainsi que les relations de citation (indiquées par l'auteur). Notons que ces inclusions par référence (appelées « transclusions »), par opposition à des inclusions par copie, permettent la citation tout en évitant le plagiat [Nelson99].

Destiné au départ au texte, le modèle a été récemment étendu aux images. Si on tentait de généraliser l'approche, on pourrait dire qu'un *document* est un ensemble ordonné de *fragments* et de *liens* réutilisables (cf. Figure 4.1).

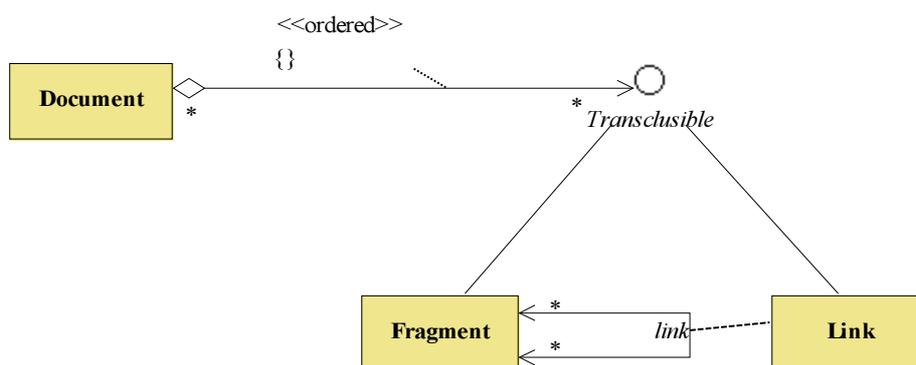


Figure 4.1 : Rétro-conception de *Xanadu*® (diagramme de classe UML)



#### CHAPITRE 4. INSTRUMENTATION DE L'INTERTEXTUALITÉ ET DE L'INTERSUBJECTIVITÉ

Notons que c'est justement cette simplicité qui a permis au « Filet » (« Mesh »), interne au CERN, de passer à l'échelle d'une « Toile Mondiale ».

Si le succès commercial et l'engouement immodéré des « internautes » pour la Toile est indubitable (au moins à l'heure où nous écrivons), il en est autrement de son usage savant. Dans un manifeste « en ligne », Ted Nelson, rêvant toujours de « machines [pour les] littéraires », décrit la Toile comme une atteinte à l'intégrité de son invention :

The Web isn't hypertext [...]. The Xanadu project has endeavored to implement a pure structure of links and facilitated re-use of content in any amounts and ways, allowing authors to concentrate on what mattered. What we have instead is the vacuous victory of typesetters over authors, and the most trivial form of hypertext that could have been imagined. [...] Instead, today's nightmarish new world is controlled by “webmasters”, tekkies unlikely to understand the niceties of text issues [...]. [Nelson02]

Dans une perspective un peu plus constructive, Michael Bieber et ses collègues [BieberEtAl97] écrivent que la Toile est aux hypermédia ce que « l'assembleur » est aux langages de programmation, c'est-à-dire seulement la « seconde génération ». Dit autrement, la Toile n'offre que des fonctionnalités de « bas niveau », difficilement utilisables pour des projets ambitieux. Par contre, l'ensemble de ces fonctionnalités est suffisant pour en construire de plus haut niveau (cf. des projets comme HyperNietzsche [IorioEtTurner99, Iorio00] ou NDWeb<sup>43</sup>). Cette idée n'était sans doute pas absente du projet originel de Tim Berners-Lee, puisque celui-ci, dès le départ, avait conçu les pages Web comme pouvant être calculées dynamiquement.

Concernant l'intertextualité, on pourra regretter à l'usage que le modèle ne permette de représenter la transclusion de plusieurs textes que par un arbre binaire dont les nœuds seraient des « cadres » et les feuilles seraient les pages. Ces « cadres » s'avèrent finalement difficilement utilisables (sans même parler de leur représentation graphique dans les navigateurs qui en a fait le cauchemar des créateurs de sites Web).

---

<sup>43</sup> <http://www.ndweb.org/recit/temple>

Pour ce qui est de l'intersubjectivité, on remarquera que si chacun peut créer une ressource subjective avec transclusion ou référence à des ressources existantes, il n'y aura aucun moyen (avec les outils standards) de remonter de la source à son commentaire : ce qui limite de beaucoup les possibilités d'interactions sociales.

### c. Hyper-G

*Hyper-G* fut conçu à l'université technologique de Graz (Autriche) par une équipe dirigée par Hermann Maurer. Le système fit l'objet de plusieurs « doctorats en sciences techniques », en particulier celui de Frank Kappe, en 1991, portant sur le serveur [Kappe91] et celui de Keith Andrews, en 1996, sur le client graphique (appelé « Harmony ») [Andrews96]. Aujourd'hui, le système est devenu un produit commercial de gestion de connaissance appelé « HyperWave »<sup>44</sup> [Maurer96].

*Hyper-G* se distingue de la Toile en gérant (cf. Figure 4.3) des *liens* bidirectionnels dont la cohérence est maintenue (absence « d'erreur 404 »), ainsi que par des primitives hypermédia de plus haut niveau appelées « *collection* ». Une collection permet de contenir plusieurs *documents* (*documents atomiques* ou autres collections). Un même document pouvant appartenir à plusieurs collections, il est possible de représenter une structure non hiérarchique. Précisons également qu'il existe deux types de collections, à savoir la *séquence* et la *grappe*, suivant que la collection est ordonnée ou pas. Notons que le document atomique porte bien son nom puisqu'il est impossible de réaliser une transclusion d'une partie de cet objet. Ceci est d'autant plus regrettable que la définition de liens, par contre, utilise des parties de documents appelées « *ancres* ».

---

<sup>44</sup> <http://www.hyperwave.com>

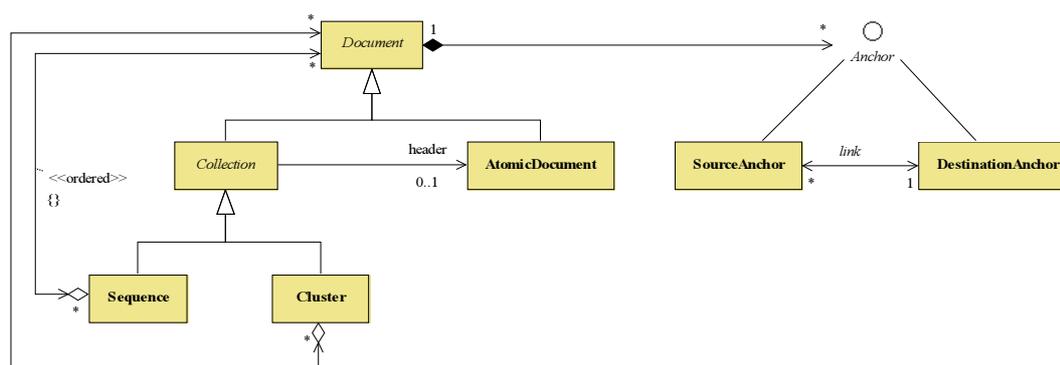


Figure 4.3 : Rétro-conception d'*Hyper-G* (diagramme de classe UML)

#### d. ATLAS.ti®

Le système *ATLAS.ti*® est issu d'un projet de recherche du département de psychologie de l'Université Technique de Berlin (1989-1992). Depuis 1993, il est commercialisé par son auteur, Thomas Muhr, en tant qu'atelier d'analyse qualitative de documents [Muhr97].

Passé pratiquement inaperçu dans notre discipline<sup>45</sup>, le modèle d'*ATLAS.ti*® n'en demeure pas moins intéressant (cf. Figure 4.4). Son élément central est la citation, fragment défini par le lecteur sur un *document primaire*. Chaque citation peut être reliée à d'autres par l'usage d'*hyperliens*, et décrite par des *codes*. Ces codes, communs à plusieurs citations, peuvent être reliés à d'autres par des liens typés (cause, équivalence, généralisation, ...). Un autre objet, le *mémo*, est un petit texte permettant de commenter un code, une citation ou un document primaire. Mémos, documents primaires, et codes peuvent être regroupés dans plusieurs *familles*. Enfin, le *supercode* se distingue du code par une définition en intension (en fonction d'autres codes, supercodes ou familles) des citations qu'il décrit.

<sup>45</sup> Merci à Ioannis Kanellos de nous l'avoir signalé.

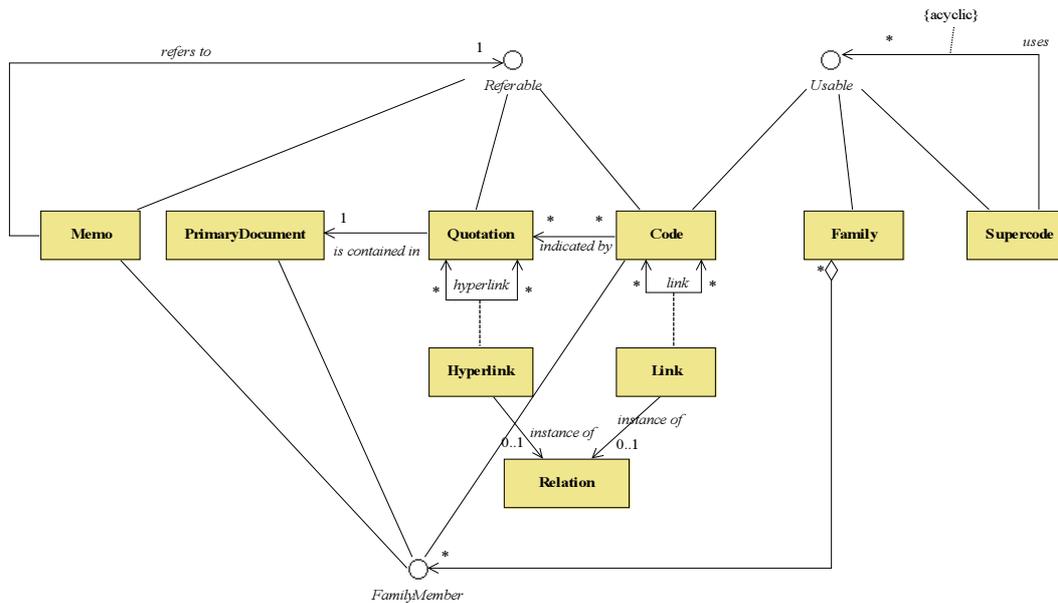


Figure 4.4 : Rétro-conception d'*ATLAS.ti*® (diagramme de classe UML)

Le modèle intertextuel présenté ici est le fruit d'une excellente connaissance de la pratique des chercheurs en Sciences Humaines. Il souffre cependant de quelques défauts de conception, notamment au niveau des généralisations. Ceci rend impossible un certain nombre d'actions qui sembleraient assez naturelles (coder un document primaire, commenter une famille ou un supercode, regrouper en une famille des citations ou des supercodes...) et complexifie inutilement le modèle.

Même s'il est possible de modéliser plusieurs points de vue en autant de « familles », on ne peut réellement parler ici d'intersubjectivité. En effet, la nature monoposte de l'application n'est guère propice au débat.

#### e. PASTEL

L'application *PASTEL* fut développée dans le cadre de la thèse de Ludovic Tanguy [Tanguy97a], thèse encadrée par Ioannis Kanellos et soutenue en 1997 à l'ENST de Bretagne.

Ce « Programme d'Aide à l'Analyse Sémantique de Textes, même Littéraires » s'appuie sur un modèle informatique [Pri95, TanguyEtThlivitis96] de la sémantique

CHAPITRE 4. INSTRUMENTATION DE L'INTERTEXTUALITÉ ET DE L'INTERSUBJECTIVITÉ interprétative de François Rastier (cf. Figure 4.5). L'interprète humain peut indiquer dans le texte à analyser un certain nombre de *sémèmes*. Les *taxèmes*, permettent de regrouper en classes plusieurs de ces *sémèmes*, chaque *sémème* ne pouvant appartenir qu'à un seul *taxème*. Les *spécèmes*, pour leur part, permettent d'opposer deux à deux les *sémèmes*. Enfin, les *sèmes* permettent de qualifier *spécèmes* et *taxèmes*, ainsi que d'entrer en relation d'*afférence* avec les *sémèmes*.

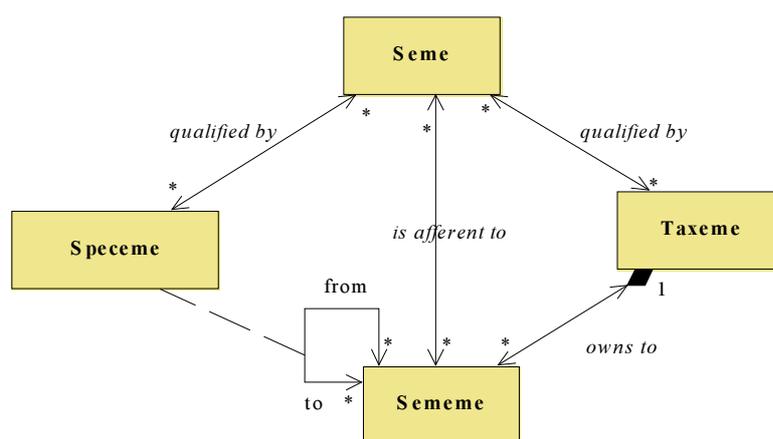


Figure 4.5: Rétro-conception de *PASTEL* (diagramme de classe UML)

De par son origine, ce modèle intertextuel est particulièrement approprié à l'analyse d'un texte. Par contre, tel quel, il est difficilement utilisable pour modéliser les différentes structures d'un corpus (en considérant que toutes sont « sémantiques »). En effet, le choix d'une sémantique purement différentielle, nous empêche de définir des relations générales (entre sèmes). Par exemple, au lieu d'indiquer une fois pour toute que la *Chronique des fouilles* se situe dans le *Bulletin de Correspondance Hellénique*, on serait obligé de préciser pour chaque passage du corpus<sup>46</sup> qu'il se situe dans les deux. Nous risquons donc d'obtenir, pour un corpus conséquent, un réseau aussi verbeux que celui de *Xanadu*®.

<sup>46</sup> Plusieurs dizaines de milliers.

## f. Strates-IA

Le modèle *Strates-IA* est issu de la thèse de Yannick Prié [Prié99], thèse préparée au LISI sous la direction d'Alain Mille et de Jean-Marie Pinon, et soutenue en 1999.

Ce modèle (cf. Figure 4.6) est dédié à l'exploitation contextuelle d'*unités audiovisuelles* (fragments de *flux audiovisuels*). Le vocabulaire de l'annotation est défini par des *éléments d'annotation abstraits* reliés par des *relations conceptuelles*. Chaque relation conceptuelle peut éventuellement être déclarée comme instance d'un *type de relation*. L'instance d'un élément d'annotation abstrait dans une unité audiovisuelle est appelée « *élément d'annotation* ». Entre deux éléments d'annotation, il est possible de définir un lien (orienté) nommé « *relation élémentaire* ».

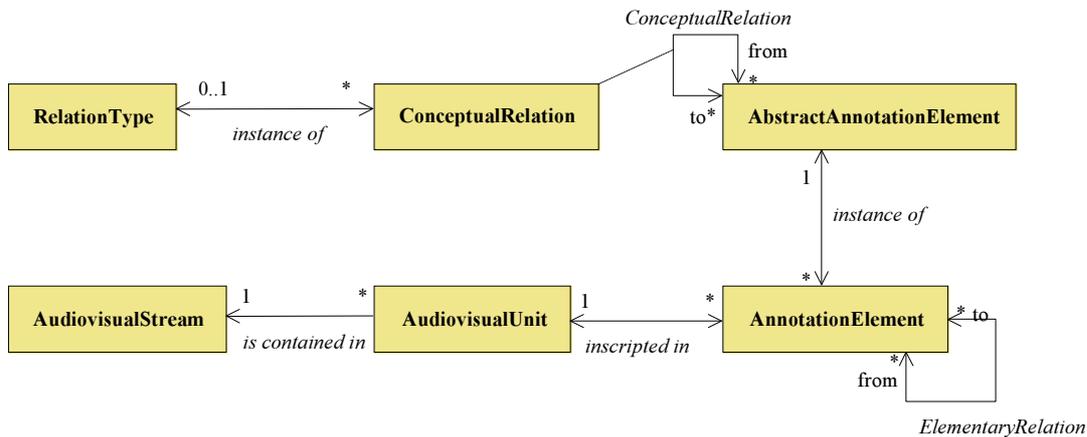


Figure 4.6: Rétro-conception de *Strates-IA* (diagramme de classe UML)

Si le modèle *Strates-IA* constitue un réseau intertextuel très réussi, l'expression de l'intersubjectivité sera par contre plus problématique. Comme l'équipe de *Strates-IA* et la nôtre l'avons expliqué dans un article commun [BenelEtAl01a], le rapport de l'individu à la connaissance diffère totalement dans nos domaines d'application respectifs. En effet, *Strates-IA* étant destiné à des documentalistes, il est naturel de séparer les autorités (éléments d'annotation abstraits) des indexations elles-mêmes (éléments d'annotation). Dans notre perspective par contre, l'expert est amené à modifier les deux de conserve. En d'autres mots, il n'est pas question qu'il suive *le* modèle du domaine mais plutôt qu'il propose *son* modèle.

Notons que l'extension de *Strates-IA* réalisée par Elöd Egyed-Zsigmond dans le cadre de sa thèse, prend en compte les notions (définies informellement dans la thèse de Yannick Prié) de *dimension d'analyse* et de *schéma de description*. Ces deux notions peuvent être utilisées pour définir différents canons d'indexation. Si ceci est un premier pas vers la notion de point de vue, il s'agit tout de même de points de vue normés par une communauté et devant être appliqués par les individus. Il ne s'agit donc pas à proprement parler d'intersubjectivité.

### g. TheBrain®

Le concept de *TheBrain*® [TheBrain01] fut inventé en 1994 par Harlan Hugh, un autodidacte en informatique alors âgé de 19 ans. La société<sup>47</sup> dont il est aujourd'hui président commercialise le produit en tant que « plateforme de connaissance ». Ce produit est protégé par une vingtaine de brevets (américains, autrichiens, italiens et européens).

Le modèle de *TheBrain*® est à la fois simple et général. Il est basé sur trois notions : celle de *contenu*, celle de *pensée* et celle de *relation* (cf. Figure 4.7). A chaque pensée, il est possible d'associer un contenu. Les pensées sont associées entre-elles par des relations de *paternité/filiation* ou de *saut*.

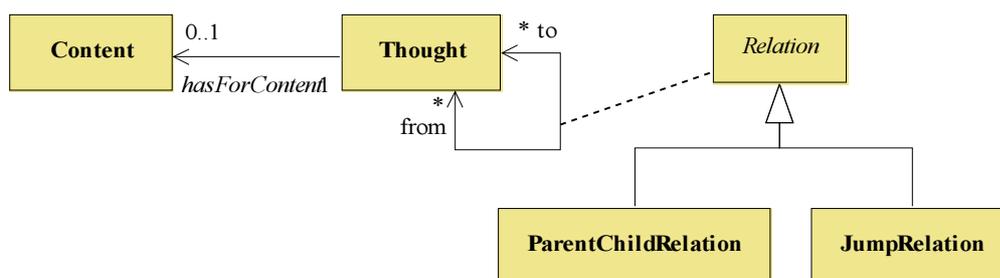


Figure 4.7 : Rétro-conception de *TheBrain*® (diagramme de classe UML)

<sup>47</sup> <http://www.thebrain.com>

La contre partie de la simplicité du modèle intertextuel réside dans un certain nombre de limites. La première concerne la gestion des contenus : ceux-ci sont gérés de manière monolithique, c'est-à-dire sans notion de fragment. La seconde concerne les traitements appliqués au réseau. Ceux-ci se bornent à l'affichage, pour un nœud sélectionné, de ses voisins (parents, enfants, frères, sauts). En effet, il serait difficile d'obtenir des affichages plus synthétiques puisque aucun contrôle de cohérence n'est appliqué au réseau. Il risquerait d'y avoir, par exemple, des paradoxes (existence d'un cycle) dans les relations parents/enfants.

*TheBrain*® (dans sa version « entreprise ») permet l'écriture collaborative d'un réseau. Cependant, l'espace partagé est ici un espace de consensus et non de débat. Nous sommes donc loin d'une instrumentation de l'intersubjectivité.

### **3. Bilan provisoire**

Suite à l'étude théorique réalisée dans les chapitres précédents, nous avons vu dans celui-ci que les enquêtes effectuées dans les bibliothèques traditionnelles corroboraient la nécessité de l'intertextualité et de l'intersubjectivité. Nous nous sommes alors intéressé à l'instrumentation de ces deux notions dans sept systèmes hypermédias (issus du monde académique ou de celui de l'industrie) qui nous ont semblé représentatifs. Il en ressort que même si l'intertextualité est un fondement de l'hypermédia, les modèles proposés sont rarement utilisables pour le cas qui nous occupe. Plus encore, la question de l'intersubjectivité reste encore largement à explorer, ne serait-ce que dans sa forme la plus simple : l'expression en un même lieu de plusieurs points de vue.

Dans de telles circonstances, l'instrumentation que nous cherchons de l'intertextualité et de l'intersubjectivité passe par la définition d'un nouveau modèle hypermédia. C'est ce que nous tâcherons de faire dans notre deuxième partie.

## **2<sup>ème</sup> partie : Modèle proposé**

