

Annotations multi-documents de points de vue exprimés

Samuel Gesche¹, Guy Caplat², Sylvie Calabretto¹

¹ LIRIS, INSA de LYON. 69621 Villeurbanne Cedex – France

{Samuel.Gesche, Sylvie.Calabretto}@insa-lyon.fr

² Département Informatique de l'INSA de Lyon – 69621 Villeurbanne Cedex – France

Guy.Caplat@insa-lyon.fr

Résumé : Nos travaux portent sur une méthodologie et une instrumentation de la confrontation des points de vue en Sciences Humaines et Sociales. Ces points de vue sont exprimés sous forme de structures hypermédia représentées dans un formalisme à base de graphes. Les résultats des confrontations, sous la forme d'annotations multi-documents, peuvent à leur tour devenir des sujets d'étude et mener à la création de nouveaux points de vue qui viendront alimenter de manière itérative le système.

Mots-clés : annotation, point de vue-opinion, confrontation

1. INTRODUCTION

Nos travaux portent sur la confrontation de points de vue dans le cadre d'un projet interdisciplinaire visant à instrumenter par l'informatique des chercheurs en Sciences Humaines. Ainsi, une plateforme nommée Porphyry [Ben02] a été créée il y a quelques années et est actuellement en exploitation. Cette plateforme permet d'ores et déjà la construction de tels points de vue.

L'objectif de ce projet interdisciplinaire est double. D'une part, il s'agit de fournir aux chercheurs des outils adaptés à leur méthode scientifique, de manière les aider au plus proche de leurs besoins. De l'autre côté, il s'agit, plus encore, de fournir des outils adaptables, afin de permettre l'évolution de la discipline scientifique et l'élaboration de nouvelles méthodes de travail par les chercheurs [Ben06].

D'un point de vue informatique, la plateforme Porphyry permet de classer des documents numériques – ou des parties de ces documents – au sein d'une hiérarchie, au moyen d'un formalisme appelé les Réseaux de Description. L'architecture est multi-tiers [Ben02], et comporte notamment un serveur de contenu pour les documents et un serveur de correspondance pour les structures.

Nous travaillons sur la confrontation de ces structures, au sein d'une extension de Porphyry nommée Platon. Cette confrontation nous a amené à réfléchir sur la notion d'annotation dans le contexte hypermédia qui est le nôtre, et nous exposerons ici le résultat de ces réflexions.

Nous parlerons donc d'abord de la notion de point de vue-opinion (ce sont, entre autres, les structures dont nous parlons, et le résultat de la confrontation de ces structures). Puis nous évoquerons rapidement les principes et objectifs de la confrontation de ces points de vue-opinion, avant de nous pencher sur les implications de ces deux notions sur la notion d'annotation, en introduisant le concept d'annotation multi-documents.

2. POINTS DE VUE-OPINION

Le système de Porphyry, auquel Platon rajoute la fonctionnalité de confrontation, est basé sur la notion de point de vue. Or le terme de point de vue, s'il est pertinent sur le plan des Sciences Humaines, pose dans le domaine informatique un bon nombre de problèmes.

2.1. Insuffisance de la notion de point de vue

En informatique, le terme de point de vue a une définition principale et une définition marginale. Commençons par la définition principale.

Comme un point de vue en montagne, on peut voir le point de vue comme un angle de vision sur quelque chose – un objet, un problème ou tout autre chose que l'on essaie d'appréhender dans l'application qu'on veut programmer – cet angle de vision étant défini par une logique de métier.

Par exemple, sur le même objet de voiture que l'on veut mettre dans une base de données, on voudra fournir une base de données unifiée pouvant servir à la fois pour un carrossier, un gestionnaire de locations, un mécanicien, un électronicien et un vendeur. Pour chacune de ces personnes, le terme « voiture » fait référence à une chose différente, mais on sait que c'est du même objet qu'ils parlent. Simplement, chacun d'entre eux s'intéresse à certaines de ses caractéristiques en ignorant les autres.

A ce moment-là, l'objectif de l'architecte de l'application est de concevoir un système qui montrera à chacun ce qu'il veut voir, et pas le reste. Ces considérations peuvent se résumer à créer des vues sur la base de données s'il ne s'agit que de consultation. Mais s'il s'agit d'un logiciel d'entreprise plus complexe, par exemple de la classe des ERP, il faudra également entrer et modifier des données. A ce moment-là, l'application devient tellement complexe qu'il est nécessaire de s'occuper de ces préoccupations aussitôt que possible dans le processus de développement. Ainsi, les méthodes de conception orientées point de vue ont fleuri ces dernières années ([Lah05], [EIA05]), et continuent de se perfectionner.

Cependant, les points de vue, dans ce contexte, sont indépendants les uns des autres. On ne va pas confronter la manière dont un mécanicien et un commercial travaillent et voient les choses. C'est même considéré, dans certains cas, comme une faute grave de management. On peut informer les personnes de ce que l'autre voit. On peut contrôler la cohérence de l'application quant aux données qui sont entrées. Mais on ne peut pas « faire discuter » les modèles respectifs.

Autre chose, ces modèles sont purement statiques. On les prévoit dans le cahier des charges, on les affine en spécifications, on les boucle en conception, mais idéalement on n'y touche plus ensuite. Dans ce contexte, doter l'outil de conception d'un contrôle de cohérence est tout ce qu'on peut faire qui se rapproche de la confrontation. Au contraire, les modèles de Porphyry sont totalement dynamiques, Porphyry permet de les construire et c'est tout.

La conclusion est donc sans appel : Porphyry ne travaille pas avec cette notion de point de vue. Cela pose problème quant à l'appellation car cette définition restera vraisemblablement la principale définition, ne serait-ce que parce qu'elle trouve davantage d'applications économiques.

2.2. Pertinence de la notion d'opinion

L'autre notion que l'on rattache au point de vue est plutôt reprise dans ce qui a été baptisé le « Web socio-sémantique » [Cah04]. Web social, parce que la discussion y joue un rôle majeur, et Web sémantique, parce que la construction de sens en est l'enjeu. Autant la notion précédente donnait lieu à des structures statiques et figées, autant celle-ci donne lieu à des structures évolutives. Pour ainsi dire, le projet est non pas de construire une application à partir de points de vue, mais de construire une application permettant de construire des points de vue. Quitte à, ensuite, construire une autre application à partir de ces points de vue s'ils sont suffisamment figés.

Dans ce contexte, on divise les points de vue en deux catégories : d'abord les points de vue « consensuels », ensuite les points de vue « opinions ». Les opinions sont ici des points de vue qui engagent celui ou ceux qui les partagent et non pas l'ensemble d'une communauté de référence. Les consensus, au contraire, sont partagés par toute la communauté. [Rib02].

C'est dans le contexte du point de vue « opinion » que Porphyry se place, et que Platon peut servir. En effet, ce qui est exprimé dans Porphyry est le résultat d'un travail de recherche. Ce travail a réuni des connaissances, des documents et des compétences pour fournir une théorie scientifique. Cette théorie peut être comparée et confrontée à d'autres théories, mais ne relève pas du consensus – et encore moins de l'angle de vision statique dont nous parlions avant.

Cette définition de l'opinion convient bien aux structures construites par les utilisateurs de Porphyry : une structure dynamique et évolutive, fruit d'un travail qui peut être attribué à une personne, physique ou morale. Cependant, dans le domaine des Sciences Humaines, l'opinion fait référence à autre chose.

2.3. Limites de la notion d'opinion

En philosophie, l'opinion est un avis que l'on considère comme vrai. La caractéristique principale de cet avis est son immédiateté : il n'y a aucun raisonnement derrière (ou un raisonnement fait une fois pour toute et probablement oublié depuis). Une opinion est donc affirmée et non élaborée. Qui plus est, elle provient souvent d'une source extérieure à la personne.

La fonction de l'opinion est à ce moment-là davantage sociale qu'intellectuelle : il s'agit de rentrer dans un groupe d'individus (ceux qui partagent la même opinion) et de se démarquer d'autres groupes (ceux qui ne la partagent pas).

En conséquence, ce concept d'opinion est dangereux pour notre définition du point de vue. En effet, les « points de vue d'experts » dont nous nous occupons sont le résultat d'un raisonnement méthodique, selon une discipline scientifique, dont nous conservons la trace. C'est exactement l'inverse.

2.4. Conclusion : le point de vue-opinion

Nous avons donc vu que le terme de point de vue est ambigu en informatique, et celui d'opinion dangereux en Sciences Humaines. D'autres termes ont été envisagés pour décrire ce qui est construit dans Porphyry (comme avis, expertise ou théorie), mais d'une part ils ont chacun un sens qui leur donne exactement la même propension à l'ambiguïté, ensuite aucun d'eux ne décrit mieux que point de vue ou opinion ce dont on parle, et enfin il est difficile de les raccrocher aux Sciences Humaines et à l'informatique. Nous avons donc préféré opter pour la combinaison de ces deux premiers termes.

En effet, le terme de « point de vue-opinion » en informatique est un synonyme d'« opinion », donc convient aussi bien. Pour les Sciences Humaines, c'est un terme a priori assez technique pour ne pas véhiculer la même signification que le simple terme d'« opinion ».

Il est à noter que le terme de « point de vue-opinion » est une contraction de « point de vue de type opinion ». Au pluriel, on parlera donc de « points de vue-opinion », opinion restant au singulier.

De manière générique (voir figure 1 ci-dessous), un point de vue-opinion est une abstraction d'un sujet d'études composé (ou non) d'un certain nombre de cas. Le point de vue

est exprimé par un modèle (au sens large), et le sujet d'études est explicité par un corpus (un ensemble de documents).

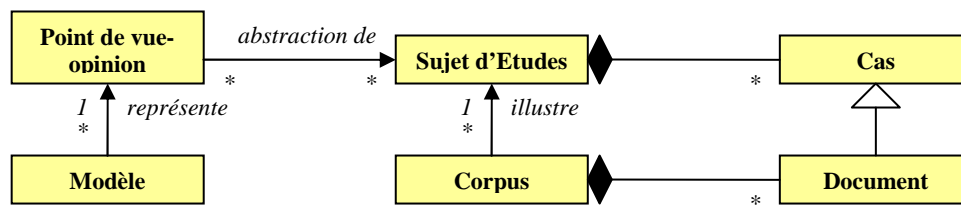


Figure 1. Modèle générique du point de vue-opinion

3. CONFRONTATION

3.1. Principe

Le principe de la confrontation des points de vue-opinion est de trouver les divergences entre leurs auteurs. Ces différences sont principalement de deux types :

- Soit l'un des modèles pointe sur des fragments documentaires que l'autre n'a pas pris en considération. Dans ce cas, il est intéressant de savoir si les points de vue-opinion couvrent le même domaine, et le cas échéant à quel point la prise en compte de ces ressources supplémentaires pourrait enrichir – ou remettre en question – le second point de vue-opinion. En effet, soit elles ont été oubliées – ou méconnues –, soit elles ont été volontairement omises (ce qui n'apparaît pas dans les Réseaux de Description de Porphyry mais peut apparaître dans un autre formalisme).
- Ou bien, la manière d'agencer les ressources et leurs pointeurs est différente, bien que les ressources elles-mêmes soient les mêmes. Dans ce cas, il est intéressant de détecter ces divergences en tant qu'interprétations différentes du même corpus.

Le principe général reste de faire émerger de nouvelles idées dans le cadre d'une dynamique de construction de théories selon une méthode scientifique.

3.2. Résultats

Le résultat escompté est un ensemble de régions des modèles où l'on a détecté des divergences. Soit une région qui n'est pas dans l'un des modèles, soit une région qui varie d'un modèle à l'autre.

Cependant, bien que le principal objectif de la confrontation soit de repérer de telles régions, il est tout aussi intéressant de confronter des critères plus généraux, confrontation qui peut donner des informations globales. Nous utilisons pour cela les critères que nous avons définis dans de précédents travaux [Ges06] :

- Complexité : le nombre de concepts et de relations entre eux ;
- Connectivité : qui indique dans quelle mesure ils sont liés (de « rien en commun » à « inclusion ») ;
- Cohérence : la cohérence intrinsèque à chacun des modèles, ainsi que leur correspondance ;
- Proximité culturelle : la proximité entre les auteurs, et entre leurs cultures respectives ;
- Proximité temporelle : à savoir si les opinions sont contemporaines ;
- Proximité thématique : entre les corpus respectifs.

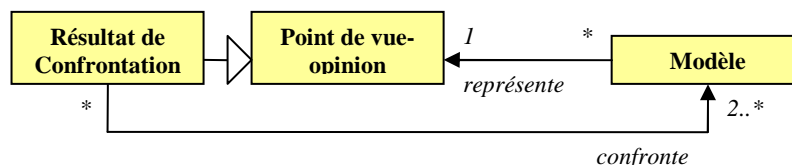


Figure 2. Modèle générique de la confrontation

De manière générique (voir figure 2 ci-dessus), un résultat de confrontation est un point de vue-opinion qui porte sur des points de vue-opinions exprimés (donc des modèles).

3.3. Formalismes

Afin d'obtenir les résultats que nous recherchons, nous envisageons la confrontation sous quatre aspects. Le premier de ces aspects est une analyse lexicale au niveau des étiquettes du graphe ; le second est une analyse syntaxique au niveau des propriétés des liens ; le troisième est une analyse protocolaire au niveau des usages éditoriaux ; le quatrième est une analyse des métadonnées, notamment en ce qui concerne les critères de confrontations généraux.

Pour que notre méthode de confrontation et les algorithmes impliqués donnent des résultats optimaux, nous devons prendre les hypothèses suivantes sur les points de vue-opinion et leurs formalismes d'expression :

- Sont concernés les formalismes de graphes uniquement. Tout autre formalisme est donc exclu de notre champ de recherche. Qui plus est, nous considérons que nous avons ou que nous pouvons construire un métamodèle de chacun de ces formalismes. Les points de vue-opinion que nous manipulons sont donc exprimés sous la forme de modèles rédigés dans de tels formalismes de graphes.
- Nous considérons également que les graphes étudiés se résument à des liens tirés entre des contenus uniquement textuels (on peut notamment remplacer les ressources non textuelles par des URI). Nous ne tenons pas compte, s'ils existent, de contenus complexes non décomposables en texte, puisque nous ne les traitons pas dans l'analyse lexicale.
- En ce qui concerne l'analyse syntaxique, nous nous basons sur la notion de propriété algébrique pour qualifier les propriétés des liens des formalismes étudiés. De ce fait, une propriété non algébrique n'est pas prise en compte (par exemple, la propriété de l'héritage qui fait que l'héritier dispose de tout le contenu de son parent).

4. ANNOTATIONS MULTI-DOCUMENTS

4.1. Annotations

[Pri00] définit l'annotation comme suit : il s'agit de l'action (ou du résultat) consistant à attacher à l'une des parties d'un document une description qui correspond à un usage que nous-mêmes – ou toute autre personne – souhaiterions en faire plus tard.

Nous définissons donc une annotation comme un ajout d'information *situé, signé et finalisé*, qui *ajoute de l'information* et cela *sans porter atteinte à l'intégrité* du texte. En d'autres termes, une annotation a une cible (un fragment documentaire), un auteur, un objectif, et tout cela en juxtaposant de l'information au texte tout en laissant la possibilité de le consulter dans son intégrité première. On peut en cela distinguer l'annotation d'un certain nombre de cas, qui sont répertoriés dans le tableau 1.

Le statut documentaire de l'annotation dépend de son adhérence au fragment de document qu'elle décrit. Dans le cas où cette adhérence est totale (par exemple gravée dans le document), l'annotation devient une partie du document au même titre que le fragment annoté. Plus exactement, l'ancien document disparaît au profit d'un nouveau document contenant le contenu de l'ancien plus l'annotation (c'est ce que l'on appelle, dans certains cas, de la dégradation d'ouvrage). Si au contraire cette adhérence est légère (comme un *post-it*) ou nulle (si elle est stockée dans une base dédiée), l'annotation peut être considérée comme un document en soi, lié au document annoté sans pour cela le modifier, comme une critique d'œuvre est liée à l'œuvre critiquée.

En informatique, les annotations sont généralement distinctes des documents annotés, et sont donc des documents à part entière [Pri05]. Ce n'est pas forcément vrai pour les annotations sémantiques [Lor06]. Les structures sémantiques sont en effet souvent rajoutées directement dans le document d'origine.

<u>Situation</u>	<u>Signature</u>	<u>Finalité</u>	<u>Ajout uniquement</u>	<u>Respect de l'original</u>	<u>Terminologie</u>
Oui	Oui	Oui	Oui	Oui	Annotation
Oui	Oui	Oui	Non	Non	Censure ou Nouvelle version
Oui	Non	Oui	Oui	Oui	Donnée d'Indexation
Non	Non	Oui	Oui	Oui	Métadonnée Documentaire
Oui/Non	Non	Non	Oui/Non	Oui/Non	Accident de Publication
Oui	Oui	Non	Non	Non	Coquille
Non	Oui	Oui	Oui/Non	Oui/Non	Trace de Réflexion
Non	Oui	Non	Oui/Non	Oui/Non	Trace d'Utilisation
Oui	Oui	Oui	Oui	Non	Correction
Oui	Oui	Non	Oui	Oui/Non	Trace de Passage
Oui/Non	Non	Oui	Oui/Non	Non	

Tableau 1 : Différentes classes d'altérations de documents

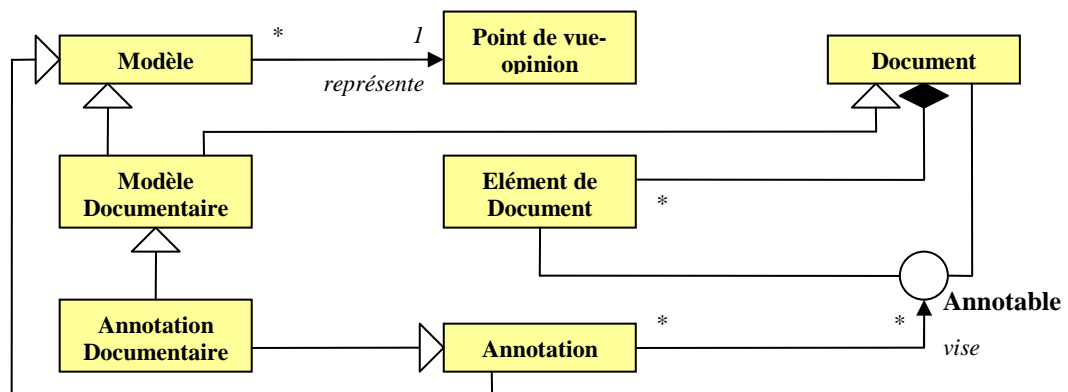


Figure 3. Modèle générique des annotations

Pour résumer (voir figure 3 ci-dessus), une annotation est un modèle vérifiant les cinq critères que nous avons donné. Ce modèle est issu d'un point de vue-opinion (qui en définit la finalité et la signature), qui porte sur une entité annotable (un élément de document ou un document par exemple). Une annotation peut être documentaire, auquel cas c'est un document à part entière. Une unique annotation peut cibler plusieurs annotables, et un unique annotable peut être la cible de plusieurs annotations.

4.2. Annotations multi-documents

[Bac99] définit le document numérique comme le résultat délivré au lecteur sur un support de communication (un écran par exemple), par opposition au fichier stocké sur une mémoire digitale (qui est la ressource servant à générer, de manière calculatoire, le document). Dans ce contexte, et avec l'avènement des hypertextes, il est possible de créer des annotations multi-documents. Par exemple, l'utilisation de la balise <iframe> dans une page HTML permet d'inclure au sein du document fourni au lecteur un morceau d'un document extérieur à la page. Si ce morceau-là contient une annotation, si plusieurs pages l'incluent, et comme chacune de ces pages est un document, on obtient bien une unique annotation partagée sur plusieurs documents.

Lorsque les annotations sont distinctes des fragments documentaires annotés, il est d'autant plus facile de définir des annotations qui portent sur un ensemble de documents, ne serait-ce que par extension des annotations qui portent sur plusieurs fragments d'un même document (un exemple est une annotation avec plusieurs flèches pointant sur plusieurs fragments distincts d'un texte).

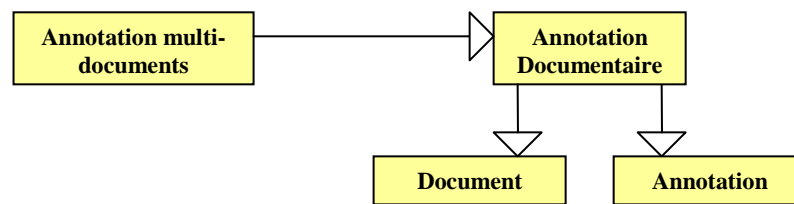


Figure 4. L'annotation multi-documents

Une annotation multi-documents est donc une annotation dont la portée (la cible) englobe plusieurs documents distincts et pas seulement plusieurs fragments d'un même document. Il s'agit forcément un document distinct des autres.

4.3. Points de Vue-Opinion

Porphyry permet de poser sur un corpus un certain nombre de structurations hypermédia. Ces structurations sont considérées comme des annotations [Ben02].

En effet, les descripteurs qui constituent les points de vue-opinion dans Porphyry sont situés puisqu'ils pointent sur des fragments documentaires précis [Ben02]. Ils sont signés parce que le point de vue est celui d'une communauté ou d'un chercheur, et non un consensus. Ils sont finalisés parce que le chercheur crée son point de vue avec un objectif bien clair, celui de sa recherche. Ils sont uniquement de l'ajout d'information et ils respectent l'intégrité des documents puisqu'ils sont indépendants des fragments documentaires, eux-mêmes étant des vues des ressources stockées dans le corpus.

Un point de vue-opinion exprimé sous Porphyry est donc bien des annotations multi-documents.

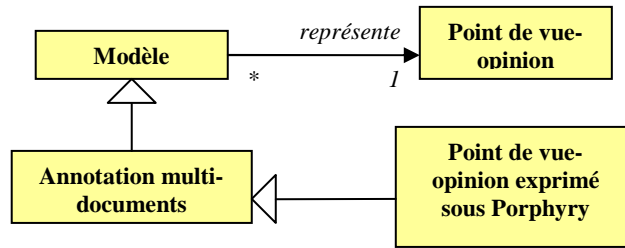


Figure 5. L'expression sous Porphyry du point de vue-opinion

4.4. Résultats de Confrontation

Platon [Ges06] permet la confrontation des points de vue-opinion construits sous Porphyry, et l'expression du résultat de cette activité de confrontation. La confrontation est guidée par la machine (en particulier au moyen de suggestions), mais la part du chercheur dans l'élaboration du résultat reste essentielle. Ce sont donc réellement des points de vue-opinion sur les points de vue-opinion que nous obtenons. Nous considérons donc ce système comme un vecteur de génération d'annotations multi-documents de points de vue-opinion.

En effet, le résultat d'une confrontation fait référence à un certain nombre de parties des points de vue-opinion en cause (il est donc situé), il est issu d'une réflexion du chercheur et non d'un processus automatique ou d'un consensus (il est signé), il est réalisé dans un objectif précis (il est finalisé), et il rajoute de l'information aux points de vue-opinion tout en garantissant leur intégrité par le fait qu'il est stocké en tant que ressource indépendante. C'est donc bien une annotation multi-documents.

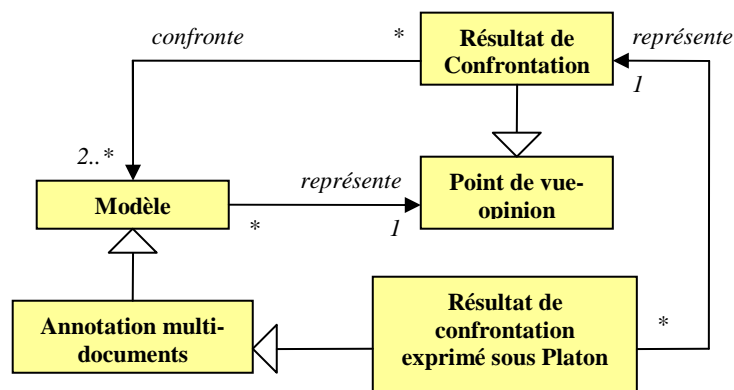


Figure 6. L'expression sous Platon du résultat de confrontation

4.5. Réinjection : des résultats comme matière première

Dans le modèle de confrontation de Platon, les résultats de confrontation sont considérés comme des documents de même niveau que les documents du corpus. C'est-à-dire qu'il est possible de les inclure dans de futurs points de vue-opinions. Les résultats de confrontation étant des annotations multi-documents et par conséquent des documents indépendants, il est possible, en effet, de les annoter de la même manière que les autres.

Nous pouvons voir dans la figure ci-dessous (figure 7) que si l'on intègre les annotations multi-documents que sont les expressions de points de vue-opinion et de résultats de confrontation, cela élargit le sujet d'étude et permet d'élaborer de nouveaux points de vue-opinion dessus.

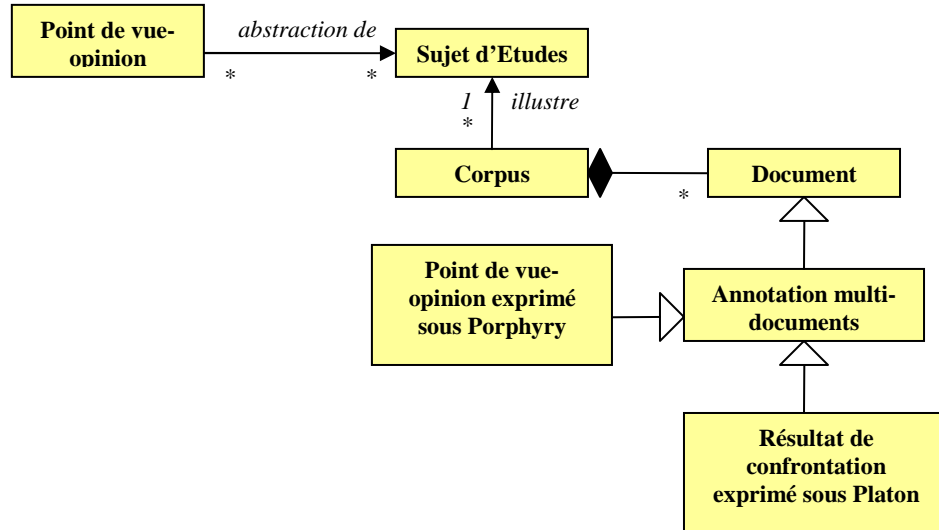


Figure 7. Réinjection des résultats dans le corpus

4.6. Modèle Général

Pour terminer, voici le modèle général qui inclut l'ensemble des modèles individuels que nous avons développés dans cette section.

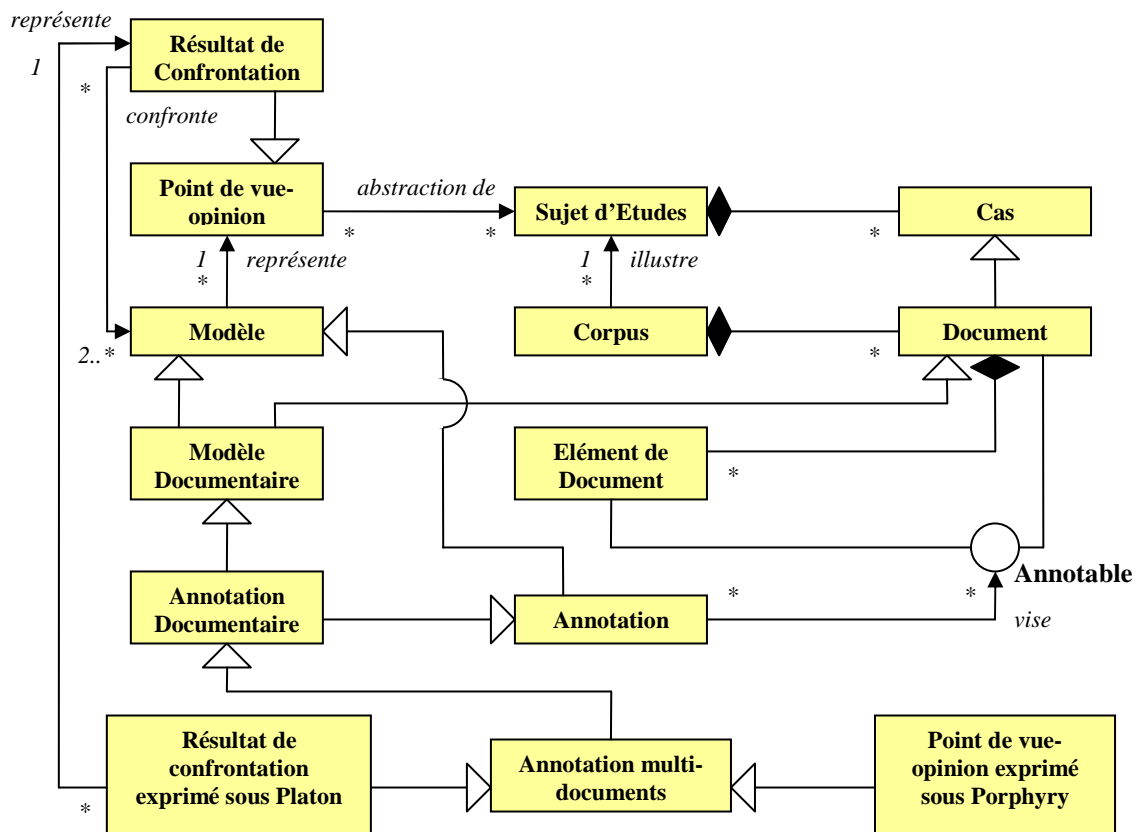


Figure 8. Modèle général

5. CONCLUSION

Nous avons présenté les notions impliquées dans le processus de confrontation de points de vue-opinions : le point de vue-opinion lui-même, issu de la différenciation à la fois d'avec le point de vue et d'avec l'opinion ; la confrontation, processus intrinsèquement humain mais donc certaines parties peuvent être automatisées ; l'annotation multi-documents, qui définit à la fois le point de vue-opinion et le résultat de confrontation par rapport aux documents qui servent à les élaborer.

6. BIBLIOGRAPHIE

- [Bac99] Bachimont B. : *Bibliothèques Numériques audiovisuelles : des enjeux scientifiques et techniques*, Document Numérique, numéro spécial « Les Bibliothèques Numériques ». Volume 2 - n°3-4/1999, pp. 319-344.
- [Ben06] Bénel A. : *Porphyry au pays des paestans : usages d'un outil d'analyse qualitative de documents par des étudiantes de maîtrise en iconographie grecque*, In: Actes du colloque international "Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation", Albi, juillet 2006. Texto!, juin 2006, vol. XI, n°2. p.182-189
- [Ben02] Bénel A., Calabretto S., Iacovella A., Pinon J-M.: *Porphyry 2001: Semantics for scholarly publications retrieval*, In. Proceedings of the thirteenth International Symposium on Methodologies for Intelligent Systems [ISMIS], Lyon, June 26-29, 2002. Lecture Notes in Artificial Intelligence, Vol. 2366, Springer Verlag. pp.351-361
- [Cah04] Cahier J-P., Zacklad M., Monceaux A. : *Une application du Web socio-sémantique à la définition d'un annuaire métier en ingénierie*, In actes de la conférence en Ingénierie des Connaissances IC 2004, Lyon, Mai 2004.
- [EIA05] El Asri B., Nassar M., Coulette B., Kriouile A.: *Multiviews Components for Information System Development*, In. Proceedings of ICEIS 2005, pp. 217-225, Miami, USA.
- [Ges06] Gesche S. : *Un modèle pour la confrontation d'opinions numérisées sous Porphyry*. In Actes du Colloque International sur le Document Electronique, CIDE'2006, Fribourg (Suisse), 18-22 septembre 2006, pp. 253-267.
- [Lah05] Lahna B., Roudies O., Giraudin J.P.: *Une approche multivue pour la conception de systèmes d'information à composants*, In actes du XVIIIème congrès INFORSID, INFORSID'2005, Grenoble, France.
- [Lor06] Lortal G. : *Annotations dans les activités coopératives : Élaboration d'un modèle générique multi-points de vue et utilisation des technologies du Web sémantique pour sa mise en œuvre*, Thèse de Doctorat. Université Technologique de Troyes, Troyes, France, 2006.
- [Pri05] Prié Y., Garlatti S. : *Métadonnées et annotations dans le Web Sémantique*, In. Charlet J., Laubet P., Reynaud C. eds. (2005) *Web sémantique*, numéro hors-série de la revue « interaction, information, intelligence ». Editions Cépaduès/Toulouse.
- [Pri00] Prié Y. : *Sur la piste de l'indexation conceptuelle de documents – Une approche par l'annotation*, Document Numérique. Volume 4 – n°1-2/2000, pp. 11-35.
- [Rib02] Ribière M., Dieng R.: *A Viewpoint Model for Cooperative Building of an Ontology*, In. Proceedings of the tenth International Conference in Conceptual Structures, Lecture Notes in Computer Science (2002), Springer-Verlag, 2393, pp. 220-234.