Lortal G., Lewkowicz M., Todirascu-Courtier A., 2005, Modélisation de l'activité d'annotation discursive pour la conception d'un collecticiel support à l'herméneutique, in *Jaulent M.-C. Actes de la conférence IC 2005* (pages 169-180). PUG, ISBN 2 7061 1284 0

Modélisation de l'activité d'annotation discursive pour la conception d'un collecticiel support à l'herméneutique

Gaëlle Lortal[1], Myriam Lewkowicz[1], Amalia Todirascu-Courtier[2]

[1] Laboratoire CNRS ISTIT, équipe Tech-CICO, Université de technologie de Troyes,
{lortal, lewkowicz}@utt.fr
[2] Université Marc Bloch de Strasbourg, ea 1339
amalia.todirascu@umb.u-strasbg.fr

Résumé : Dans un contexte où les échanges médiatisés s'accroissent, le document numérique devient central. Pour soutenir les échanges et la construction d'une interprétation collective autour de ce document, il est nécessaire de développer des collecticiels pertinents, support à l'herméneutique. Dans ce cadre, l'annotation est un fragment de discours à propos d'un texte, un support à l'argumentation. Dans cet article, il s'agit de présenter une démarche de conception d'outil fondée sur la modélisation de l'activité d'annotation discursive. Ce modèle est issu des travaux en rhétorique antique et médiévale sur la production de discours. De ce modèle d'activité d'annotation, nous dégageons des primitives de conception de l'outil support à l'herméneutique.

Mots-clés : Annotation, discours, création collective de sens, conception de collecticiel.

1 Introduction

La place omniprésente des documents dans nos organisations a donné lieu à des travaux de recherche focalisés sur le document, comme par exemple en France, les travaux au sein du Réseau Thématique Pluridisciplinaire 33 : « Documents et contenu : création, indexation, navigation » (RTP-DOC). Ce réseau distingue trois orientations de recherche associées au document : l'analyse du document comme forme (études de la structure du document pour sa manipulation), comme signe (étude de l'intentionnalité du document), et comme médium (étude du statut du document dans les relations sociales) (Pédauque, R.T., 2003). Le travail de recherche que nous présentons dans cet article s'intègre dans la seconde problématique qui considère le document « porteur de sens et doté d'une intentionnalité […] indissociable du sujet en contexte qui le construit ou le reconstruit et lui donne sens » (Pédauque, R.T., 2003, p. 3). Dans cette vision du document comme signe, on s'intéresse plus particulièrement à la création d'un document, à l'interprétation d'un document, c'est-à-dire aux signes qui le constituent. Ces questions sont abordées dans cet article sous l'angle de la lecture *critique* des documents, que nous opposons à une lecture qui ne serait guidée

par aucun principe productif, qui ne viserait ni un savoir, ni la production d'un autre texte.

Dans cet article, nous présentons tout d'abord l'annotation comme élément discursif central d'une lecture critique. Nous exposons ensuite les principes méthodologiques pour la conception d'un collecticiel support à l'herméneutique, puis faisons le point sur les travaux existants sur la modélisation des activités liées à l'écrit. Nous proposons (section 5) un modèle de l'activité de production de discours issu de la rhétorique, décliné (section 6) dans le cadre d'une activité instrumentée. Ce modèle sert de base à la conception du collecticiel dont nous décrivons les fonctions (section 7). Nous précisons enfin comment nous proposons d'utiliser des techniques issues du Traitement Automatique de la Langue (T.A.L.) afin d'appareiller le support textuel.

2 L'annotation : une production discursive de la lecture critique

Une lecture critique est productive d'une interprétation qui éclaire non seulement le texte lu, mais également d'autres textes. Elle permet de contextualiser le texte et de le faire entrer ainsi dans un contexte discursif. Cette lecture critique peut être soutenue par un ensemble de fragments textuels reliés au texte et formant son « co-texte » (contexte textuel), les annotations. Elle peut donc être productive d'un autre texte, commentaire ou critique, le corps textuel de l'annotation marquant une argumentation autour d'un texte, un point de vue d'un lecteur. Nous nous intéressons plus particulièrement à une lecture critique à plusieurs, permettant la construction d'une interprétation partagée du document initial. Cette élaboration d'une interprétation partagée au sein d'un collectif participe selon nous à la construction collective du sens (Weick, 1979). Ce dernier conçoit en effet la construction collective du sens dans les organisations (*collective sensemaking*) comme un processus de réduction collective de l'ambiguïté perçue d'une situation. C'est en échangeant, en débattant, que les membres de l'organisation vont clarifier puis partager des compréhensions de situations (retranscrites dans des documents), ce qui construira du sens petit à petit. Les travaux de Weick mettent l'accent sur le processus de création du sens, son émergence et son évolution, et non pas sur la représentation collective du sens. Le sens collectif n'est donc pas forcément un sens partagé par un collectif. L'interprétation collective de documents, traces des actions menées dans l'organisation, permet de tirer parti des documents tout en étant capable de sortir éventuellement du cadre dans lequel les documents ont été rédigés. Ce processus est également le support de l'identité individuelle, car, par le biais de ces interactions, chaque acteur met à l'épreuve et fait évoluer son identité.

Nous proposons de soutenir l'herméneutique, l'élaboration de discussions critiques ou explicatives autour de textes, en développant des stratégies d'interactions médiatisées autour des documents numériques, souvent textuels. L'herméneutique consiste à recréer un sens autour d'un texte qui ne possède plus qu'intrinsèquement son contexte. Ce sens laissé au sein du texte même par l'auteur est redécouvert par une succession d'énoncés expliquant une interprétation, se complétant, se répondant, construisant de nouvelles pistes d'interprétation. L'interprétation des textes est

traditionnellement accompagnée de gloses, commentaires et autres annotations ancrées au texte même, ou reliant différents textes ou fragments de texte. Nous proposons donc de soutenir cette collaboration discursive autour des documents par un système permettant l'annotation de ces documents, avec une finalité d'interprétation et d'appropriation, finalité non assistée par les outils informatiques d'annotations textuelles actuels, qui ne permettent de déposer que des annotations isolées, du type commentaire textuel, pauvrement indexées (date, nom d'auteur) et difficilement utilisables comme support aux interactions au sein d'un collectif. Or dans un contexte herméneutique d'interprétation méthodique de textes, le corps textuel des commentaires est placé au rang de discours, son contexte est formé notamment par le rôle de l'auteur, le contenu sémantique, la place de cette annotation dans le fil de discussion. Cette contextualisation est essentielle pour tracer la logique de conception d'une interprétation, voire d'un concept.

Des travaux ont été menés au KMI (Knowledge Media Institute) sur ces fonctions de commentaires discursifs d'un document. Ils ont donné lieu au "Digital Document Discourse Environment" (D3E) (Sumner et al., 2000), outil web dans lequel des échanges de messages « autour » d'un document peuvent avoir lieu, mais la conception de cet outil n'étant pas liée à une étude de l'activité d'analyse documentaire, il n'existe pas de réflexion sur le processus coopératif d'interprétation partagée. De plus, la réflexion sur la visualisation et la réutilisation de ces messages est peu approfondie ; en effet, les messages sont présentés de manière arborescente, et indexés selon des attributs standards (date, auteur, titre). Tout se passe comme si on avait associé un forum au document. Or, de nombreux travaux soulignent que les discussions en ligne sont souvent désorganisées et confuses, à cause du développement fréquent de multiples fils de discussion et de conversations parallèles. On peut citer par exemple (Marcoccia, 2004), qui évoque le phénomène de digression thématique à l'intérieur d'un forum, qui se fait progressivement, en parcourant une chaîne de messages introduisant chacun un développement thématique par rapport au message précédent. Le résultat peut être une véritable « décomposition thématique » (*topic decay*, Herring, 1999).

3 Principes méthodologiques pour la conception d'un collecticiel support à l'herméneutique

La situation de recherche dans laquelle nous évoluons nous amène à définir de nouvelles pratiques à assister (l'interprétation collective de documents numériques). De ce fait, un processus de conception classique en informatique basé sur une analyse des besoins, ou sur une analyse de l'activité existante, pour en déduire des primitives de conception, n'est pas adapté. Comme le dit Tchounikine : « la recherche en informatique a donc ici un rôle fondamental, celui d'inventer de nouveaux possibles » (Tchounikine, 2002-b, p.207).

La démarche que nous proposons est inspirée du positionnement méthodologique adopté dans le champ de la conception des EIAH (Environnements Informatiques pour l'Apprentissage Humain) par Baker (Baker, 2000), repris par Tchounikine (Tchounikine, 2002-a). Ces auteurs distinguent les modèles comme outil scientifique

des modèles pour la conception de systèmes. Les premiers permettent d'utiliser une théorie pour comprendre ou prédire une situation ou une activité ; les seconds traduisent les premiers en un modèle permettant la conception et l'implémentation de systèmes supports à la situation ou à l'activité.

Or les théories issues des sciences humaines habituellement mobilisées lors de la conception de collecticiels (théorie de l'activité, théorie de l'apprentissage, théorie de l'agir communicationnel…) sont difficiles à exploiter telles quelles pour en déduire des primitives de conception, ou il est difficile de transposer leurs définitions dans un cadre médiatisé par un système informatique. Le travail de conception consiste donc à définir de nouveaux modèles, avec de nouveaux concepts, en accord avec la théorie, pour décrire l'artefact assistant et traçant les interactions. La théorie nous permettra ensuite d'analyser les traces ainsi mémorisées.

Nous proposons donc la démarche suivante, illustrée en figure 1 : dans le cadre d'une théorie en sciences humaines ou sociales adaptée aux phénomènes que l'on souhaite assister/observer, nous proposons un modèle de description de ces phénomènes. Ce modèle opérationnalise la théorie et est une base de réflexion à la définition des situations dans lesquelles ces phénomènes seraient médiatisés à l'aide d'un système informatique. Cette réflexion conduit à un modèle de l'activité instrumentée, mettant en jeu à la fois les chercheurs en sciences sociales garant du modèle de description, et les chercheurs en informatique, comprenant et maîtrisant les propriétés propres des outils informatiques. Ce modèle de l'activité instrumentée est ensuite matérialisé dans un modèle de conception, spécification du collecticiel. Ce collecticiel supportant les interactions est également un moyen privilégié de recueil de corpus. Ce corpus, analysé à l'aide de la théorie mobilisée, nous permettra de faire évoluer notre compréhension des phénomènes à l'étude.

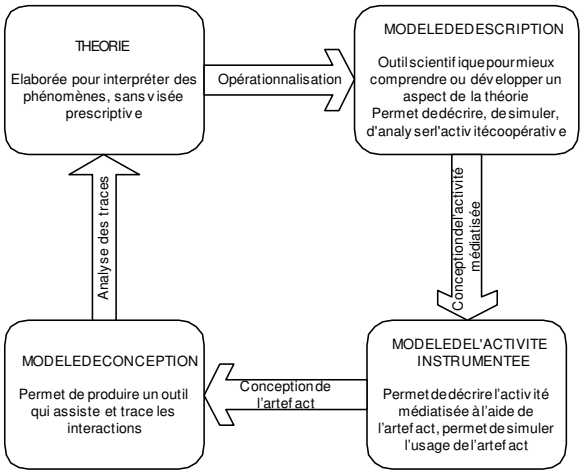

Fig. 1 - Démarche de conception d'un collecticiel

Il nous semble que bien que la phase de conception de l'activité instrumentée soit toujours présente lors de la conception d'un artefact informatique, les activités de cette phase ne sont pas explicitées. Tout se passe comme s'il était possible de définir directement des primitives de conception d'un artefact assistant une activité à partir de la modélisation du déroulement de cette activité en face à face, que cette modélisation soit réalisée par des chercheurs en sciences sociales ou non. Or nul ne nierait que cette instrumentation a un impact sur l'activité. C'est au cours de cette phase de conception de l'activité médiatisée que les échanges entre chercheurs en sciences humaines (SHS) et chercheurs en technologies de l'information et de la communication (STIC) vont pouvoir s'exprimer afin de construire un modèle commun reflétant à la fois les principes directeurs de l'activité et les possibilités en terme d'assistance. Cette phase permet le passage de relais vers des questions de conception qui seront explicitées dans un modèle de conception décrivant les fonctions de l'outil.

Dans la section suivante, afin de définir le modèle de description le plus adapté à notre problématique d'herméneutique, nous évoquons les travaux ayant analysé les activités centrées sur les documents.

4 Quelle théorie pour l'analyse de l'activité de production de discours ?

Dans le domaine de la psychologie cognitive de nombreux chercheurs ont étudié les activités mentales liées à l'écrit, en distinguant les activités de compréhension de texte et de production de texte.

En ce qui concerne les modèles de compréhension, l'accent a été mis sur la mise en mémoire de fragments de texte, fragments nécessairement résumés. Un des modèles les plus cités dans ce domaine est celui du processus de construction-intégration de (Kintsch, 1988) ; La compréhension de texte y est décrite comme un cycle alternant des phases de construction d'une représentation mentale cohérente d'un texte en cours de lecture et des phases de sélection ou non des fragments de texte pour la mise en mémoire (intégration). Des recherches ont été menées pour utiliser cette théorie descriptive à des fins constructives, par exemple pour la définition de principes de conception de documents hypermédia facilement appropriables par le lecteur (Iksal, Garlatti, 2000). Ces auteurs proposent un guide de « bonnes pratiques » pour la conception de documents, afin d'assurer notamment la cohérence du texte. Ces documents sont ensuite présentés de manière à ce que le lecteur soit assisté dans la construction de son modèle mental, l'objectif étant la minimisation du coût cognitif lors de la lecture des documents.

En ce qui concerne les modèles de production, l'accent est mis sur les processus rédactionnels de planification, de mise en texte et de révision, et le modèle de contrôle permettant d'appliquer ces processus. Les auteurs fréquemment cités dans ce domaines sont notamment (Hayes, Flower, 1980) qui ont proposé des modèles de stratégies rédactionnelles. Là encore, cette théorie descriptive a été utilisée dans des travaux qui ont donné lieu à des logiciels support aux processus rédactionnels. On peut citer par exemple les travaux de (Piolat et al., 1989) qui utilisent la combinaison

de trois logiciels (scripsis, scripap, scriprev), chacun étant focalisé sur un processus (planification, mise en texte, révision). L'objectif de ces travaux n'est toutefois pas de proposer des outils de production de texte au sein d'une organisation, mais de fournir un cadre à l'étude expérimentale de la production de texte.

Comme nous l'avons présenté en section 3, notre approche consiste à concevoir un collecticiel sur la base d'une analyse de l'activité collective que ce collecticiel entend assister. Les modèles descriptifs de compréhension ou de production proposés par la psychologie cognitive que nous avons cités ci-dessus, ne nous paraissent pas appropriés pour la conception d'un outil support à l'herméneutique car ils séparent des phases de mise en mémoire et de mise en texte. En effet, l'herméneutique mêle l'activité d'écriture durant la lecture – annotations – à celle de lecture pour produire du sens. Les phases de lecture/mémorisation et de rédaction/intégration sont donc associées. Dans des conceptions liées à la didactique de l'écrit, la lecture et la rédaction sont également considérées comme des phases d'une activité générique liée au support écrit (Barré de Miniac, 2000). Nous proposons donc de mobiliser un modèle de la production de discours issu de la rhétorique antique et médiévale représentant la mémorisation et la production discursive en un cycle complet.

5 Modèle de production de discours

L'écrit est le lieu d'interactions complexes et évolutives entre des facteurs affectifs, cognitifs et linguistiques (Barré de Miniac, 2000). Nous nous intéresserons plus particulièrement aux facteurs cognitifs en tant que facteurs organisateurs des concepts en mémoire et en texte et aux facteurs linguistiques en tant que marques à la fois d'un type de discours spécifique et de la sémantique du document en « co-texte ».

Nous retrouvons ces deux types de facteurs dans la rhétorique. Des théories rhétoriques d'Aristote à celles d'Hugues de St Victor en passant par Cicéron ou Quintilien, la production de discours est enseignée suivant un processus défini. La rhétorique aristotélicienne se focalise sur une production finale de discours oral sans nier pour autant une phase mémorielle nécessaire à toute production. Cette phase de mémorisation est mieux représentée par la rhétorique que nous appellerons mémorielle portée par des penseurs cités par (Carruthers, 1990), tels que Quintilien (*L'institution oratoire*), Cicéron (*De oratore*, *De inventione*) ou Tullius (*Ad Herennium*) dans l'Antiquité, puis Hugues de St Victor (*Didascalicon*), Fortunatianus (*Artis rhetoricae libri tres*) ou Julius Victor (*Ars rhetorica*) au Moyen-Âge. Dans cette approche de la rhétorique, un continuum entre la partie mémorielle plus « logique » ou « dialectique » et la partie stylistique, rédactionnelle, est observable. La rhétorique est considérée comme une alliance entre structuration et éloquence.

Le processus de production de discours tel qu'il est préconisé dans ce contexte comporte deux phases : « Divisio » et « Compositio ». La Divisio se fait au cours de la lecture et représente l'étape de division d'un texte en unités intelligibles, en segments brefs mémorisables. La Compositio, elle, est l'assemblage ordonné, l'agencement convenable des « res » (objets conceptuels comme physiques) des segments mémorisés (fig.2). Ces phases de mémorisation, la Divisio, et de création, la Compositio, sont elles-mêmes divisées en étapes soutenues par l'utilisation d'annotations.

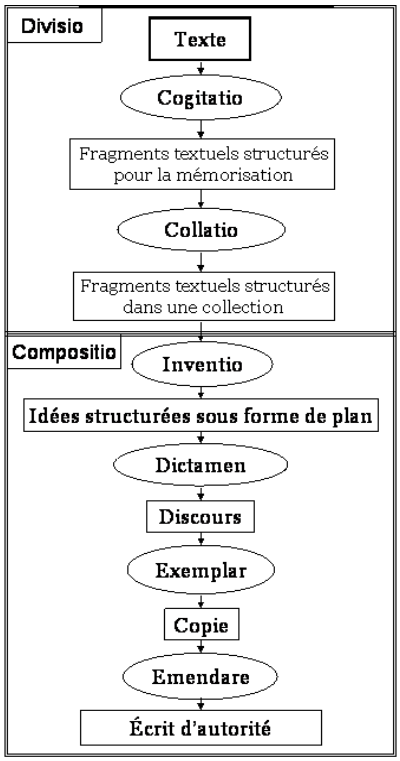

Fig. 2 : modèle de production de discours

La première étape de la Divisio est la Cogitatio. C'est une étape mémorielle individuelle qui consiste en l'association d'images par un choix et une remémoration conscients et d'une division chronologique du contenu d'un document en différents lieux mémoriels. Les fragments textuels formant le texte sont alors structurés et deviennent mémorisables facilement.

La Collatio est la phase où l'on combine en une structure des fragments textuels reposant en plusieurs lieux de mémoire distincts. C'est dans cette phase que des liens entre les différents lieux de contenu sont créés. Un « co-texte » est formé entre les nouveaux fragments mémorisés et les fragments précédemment mémorisés, en les liant sémantiquement entre eux. Cette phase n'est pas spécifiquement individuelle même si elle structure une mémoire individuelle, dans la mesure où cette étape peut être liée à des échanges discursifs, des interactions avec d'autres aidant à une structuration enrichie des concepts.

La Compositio est divisée en quatre stades d'activité évoquant des stades de création de document. Le stade d'Inventio est proche de celui de Collatio dans la mesure où il s'agit de créer des liens sémantiques entre divers éléments mémorisés, au niveau de la « res » (objets conceptuels, idée) pas au niveau du mot. Un plan est formé, c'est-à-dire un ensemble hiérarchisé d'idées, une structure argumentative par exemple.

La phase suivante sera celle de mise en mot de ce plan conceptuel, une phase classique de rédaction, le « Dictamen ». Nous voyons à cette étape la création physique du discours, classiquement sur support encore modulable (brouillon), où seul le style, le choix des termes, donc la forme textuelle du discours peut être modifiée.

La phase d'Exemplar n'est que la mise en support pérenne d'un discours strictement identique à ce que l'on trouve en sortie du processus du Dictamen.

La dernière phase et non la moindre dans cette succession de processus est celle de l'Emendare où la copie finale du discours diffusée est commentée publiquement, par l'ajout des commentaires, « notae » ou arguments d'un auteur au texte original, faisant ainsi du texte une référence, un écrit faisant autorité.

Ce modèle représente un mode de production de discours fortement soutenu par la mémoire. Dans un contexte de travail collaboratif médiatisé par ordinateur, la création discursive doit être soutenue par un outil adéquat permettant de stocker, de créer et partager les informations. Afin de concevoir cet outil, nous souhaitons donc tout d'abord modéliser cette activité de production de discours instrumentée, activitée que nous n'analysons ici que dans un cadre de documents numériques textuels.

6 Modèle de production de discours instrumentée

La déclinaison du modèle de production de discours dans un cadre instrumenté nous permet de définir les étapes suivantes à préconiser (figure 3) : tout d'abord, le texte du document est *segmenté* pour être mis en mémoire sous la forme de fragments mémorisables. Ces segments sont ensuite *indexés* pour éviter la perte de la structuration du document comme unité. Il est important d'indexer les segments chronologiquement pour marquer la hiérarchie des différents paragraphes dans un texte, des différents mots dans un paragraphe,… Ce type d'indexation concerne toutes les méta-données associables automatiquement à un élément déposé (localisation, auteur, date, …). L'indexation doit également servir à lier les nouveaux fragments déposés à l'ensemble conceptuel existant. Nous obtenons alors un ensemble de segments textuels liés sémantiquement à d'autres segments textuels. La phase de *structuration* représente un processus de hiérarchisation, d'organisation des idées, selon un plan chronologique. Un plan détaillé est défini, contenant toutes les idées nécessaires à la mise en mot du discours. C'est la phase où les « res » (concepts) contenus dans les fragments textuels indexés sont réutilisés et réorganisés en un nouveau document. La phase de *rédaction* est une mise en texte du plan donnant en résultat le discours. Ce

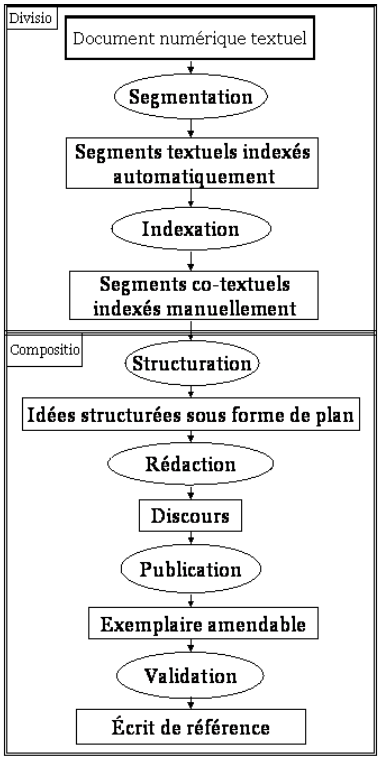

Fig. 3 – Modèle de production de discours instrumentée

discours n'est pas l'objectif final de cette activité dans cette vision de la rhétorique, puisqu'il est ensuite *publié* pour en faire un objet amendable par d'autres lecteurs/auteurs de la communauté. Cette phase de commentaire par d'autres membres de la communauté est primordiale pour permettre la *validation* de l'exemplaire, voire son amélioration, et constituer un écrit d'autorité, un discours faisant référence dans la communauté. L'annotation d'un document dans un cadre herméneutique consiste donc selon nous à suivre un processus de mise en discours d'idées organisées. Il s'agit en effet, suite à la lecture d'un document, d'engager un processus qui permet d'ajouter une idée, une opinion, structurée sous forme textuelle.

Par exemple, dans un contexte de travail collaboratif, on peut considérer la mise en partage d'un document afin qu'il soit commenté. Après une phase de visualisation du texte, d'une lecture, un segment sera mis en valeur de façon à signaler l'ancrage d'un élément discursif en rapport avec ce segment. Cette mise en valeur pourra se faire par des techniques classiques de surlignage, soulignement, encadrement, colorisation de segments de tailles variables (du mot, ou d'une partie d'un mot, au paragraphe, ou d'un ensemble disjoint d'éléments). Suite à cette segmentation et ce choix d'élément(s) à annoter, une phase d'indexation pourra intervenir, consistant en une mise en relation des segments. Il sera alors question de tirer des liens sémantiques entre des éléments pour les structurer entre eux et former un ensemble de segments textuels organisés selon leur sens, donné par l'utilisateur. L'annotation consistera en effet en un ancrage, une relation géographique, mais aussi en un corps, un discours qui fait sens, qui prend place dans un « co-texte », l'ensemble des segments textuels en mémoire indexés par des mots-clés compréhensibles et structurants pour un utilisateur humain. Lors de la rédaction de cette annotation, une phase d'organisation du discours à écrire sera nécessaire, c'est la structuration des « res », des concepts en mémoire, qui donnera naissance à un plan constitué d'arguments structurés hiérarchiquement. La phase de rédaction permettra de constituer le corps de l'annotation qui sera lisible par un membre de la discussion herméneutique après sa publication et donc sa diffusion.

Tout comme un texte de référence, l'annotation peut être entérinée grâce à un nouveau lien amené à celle-ci. Une réponse à un commentaire permet alors de participer au fil de discussion initié par la première annotation. Le passage du commentaire individuel à l'annotation discursive marquant la naissance d'une argumentation autour d'un document s'articule grâce à un modèle de coopération sous-jacent représentant les interactions créées par l'utilisation d'annotation. Ce modèle est en cours d'élaboration et n'est donc pas présenté ici.

Le modèle de production de discours instrumentée présenté dans cette section est issu d'un modèle d'activité de production de discours venant de la rhétorique. Il nous permet dans un premier temps de représenter la rédaction individuelle d'annotation et dans un second temps de décrire les spécifications d'un collecticiel assistant ce type de production discursive par le biais d'annotation.

7 Pistes de réflexion pour la conception d'un collecticiel support à l'herméneutique

Conformément au modèle que nous préconisons, le collecticiel doit permettre aux utilisateurs de visualiser un document, de le segmenter, de créer des associations de divers types (indexation, assemblage) entre les différents fragments, de rédiger le discours du corps de l'annotation, ou de la publier. La phase de validation, optimisant la collaboration au travers des réponses à un discours, nécessite la mise en place d'une fonction d'association spécifique de type « réponse à » une annotation. Le modèle discursif devant permettre un aller-retour constant entre lecture et écriture, la fonctionnalité de visualisation est également prépondérante. Entre autres pour cet objectif, l'utilisation d'un plug-in dans un navigateur semble pertinente. En effet, n'étant qu'un ajout à un navigateur naturellement utilisé et donnant accès en lecture à de nombreux documents du Web, ce plug-in permet de visualiser à la fois le document et le corps de l'annotation en cours de rédaction, ou en cours d'indexation. Cette annotation est saisie dans une fenêtre « pop-up », puis indexée pour autoriser sa récupération après publication et la création d'un ensemble de documents structurés.

Nous proposons d'adopter une architecture client-serveur respectant le standard Annotea du W3C (Kahan et al., 2001), pour la gestion des annotations et le développement d'outils d'annotation, basé sur une description des annotations en RDF améliorant la collaboration au travers de méta-données partagées. Le serveur d'annotations Zannot respectant ce standard (Zannot, 2003) conserve les annotations dans une base de données RDF et les utilisateurs peuvent interagir avec le serveur par le client Annozilla (Annozilla, 2004) afin de rechercher, créer ou supprimer une annotation. Une annotation suit une notation en RDF qu'il est possible de personnaliser (ajout de valeurs d'attributs au schéma d'annotation Annotea), ce qui permet d'ajuster le modèle en fonction de notre besoin. Nous avons choisi de réutiliser le plug-in de Mozilla nommé Annozilla qui fournit l'interface de gestion de l'annotation. Nous proposons de l'augmenter avec des fonctionnalités plus précises d'indexation de l'annotation et de visualisation des annotations en fonction des critères d'indexation.

L'indexation de l'annotation peut se faire automatiquement par l'outil (date, auteur, codifications des annotations répondues, fil de discussion chronologique automatique) ou manuellement par l'utilisateur, selon trois dimensions : domaine, argumentation, organisation (Zacklad et al., 2003). Dans le second cas (choix d'une valeur représentant le contenu de l'annotation pour chacune des dimensions), l'indexation peut être fastidieuse et nous souhaitons donc soutenir l'utilisateur dans cette tâche en lui proposant des termes (mots-clés) automatiquement grâce à des outils de Traitement Automatique de la Langue (T.A.L.). Pour cette opération, nous souhaitons utiliser un algorithme de mise en correspondance entre le corps de l'annotation et trois ontologies dimensionnelles semi-formelles. Le système d'annotation proposera alors à l'utilisateur des mots-clés pour chacune des dimensions. L'utilisateur décidera ensuite si l'indexation proposée est pertinente et s'il faut alors la conserver comme méta-donnée de son annotation. Une fois la validation effectuée par l'utilisateur, l'annotation est stockée avec ses méta-données sur le serveur d'annotations.

8 Conclusion, perspectives

Dans un environnement où le document numérique est omniprésent, nous nous sommes penchés sur une problématique spécifique de création de sens, l'interprétation collective de document, appelée herméneutique dans la tradition des textes, et soutenue par l'activité d'annotation. L'annotation est un medium privilégié d'interprétation collective engageant une communication dans un collectif, un fil de discussion. Afin de soutenir cette activité dans un cadre médiatisé, nous proposons de concevoir un collecticiel support au discours via les annotations. Les fonctionnalités de celui-ci se fondent sur un modèle d'activité de production de discours représentant les différentes étapes de production d'une annotation discursive. Nous proposons pour cet outil une architecture respectant le standard Annotea du W3C et des fonctionnalités soutenues par des techniques de T.A.L.

Dans le cadre d'une conception itérative du collecticiel, une maquette de l'outil a montré la faisabilité de l'approche. Finalisée, elle permettra une évaluation de nos hypothèses sur le modèle de production de discours, mais aussi sur le statut et les objectifs de l'annotation.

En parallèle, une expérience sur le processus d'annotation dans un groupe de conception est en cours et les données recueillies permettront d'avancer sur la définition des dimensions argumentative et organisationnelle indexant les annotations.

Les annotations seront observées non seulement pour affiner une typologie des annotations dans le cadre d'activités de conception collaborative, mais aussi pour proposer un modèle du processus d'interaction au travers des annotations déterminant les fonctionnalités coopératives de l'outil.

Références

ANNOZILLA (2004), http://annozilla.mozdev.org/

BAKER M. (2000). The roles of models in Artificial Intelligence and Education Research: a prospective view. *International Journal of Artificial Intelligence in Education Research.* Vol 11(2), p. 122-143.

BARRÉ DE MINIAC C. (2000) *Le rapport à l'écriture : Aspects théoriques et didactiques* coll. Savoirs mieux Ed. Septentrion Presses Universitaires, Ch. Barré de Miniac.

BUITELAAR P., OLEJNIK D., HUTANU M., SCHUTZ A., DECLERCK T., ET SINTEK, M. (2004), *Towards Ontology Engineering Based on Linguistic Analysis*, in Proceedings of LREC'2004, Lisbon, may 2004, ISBN 2-9517408-1-6, pp.7-11

CARRUTHERS M. (1990) *The Book of Memory: A Study of Memory in Medieval Culture*. New York: Cambridge University Press.

CIMIANO, P. HOTHO, A., ET STAAB S. (2004), *Clustering Concept Hierarchies from Text*, in Proceedings of LREC'2004, Lisbon, may 2004, ISBN 2-9517408-1-6, pp. 1721-1724.

FAYOL M. (1997). *Des idées au texte : psychologie cognitive de la production verbale, orale et écrite*. Paris: PUF.

GARLATTI S., IKSAL S. (2000). Méthodologie de conception de documents électroniques adaptifs sur le Web. IN . GAIO, M., TRUPINCIDE, E., *Document Électronique Dynamique, Actes du troisième colloque international sur le document électronique : CIDE'2000*.

HAYES J. R. & FLOWER, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing.* Hillsdale, NJ: Lawrence Erlbaum.

HERRING, S.C. (1999). Interactional Coherence in CMC. *Journal of Computer-Mediated Communication* 4(4) : www.ascusc.org/jcmc/vol4/issue4/

JACQUEMIN C. ET BOURIGAULT D. (2003), Term Extraction and Automatic Indexing, in Mitkov R. (ed), *The Oxford Handbook of Computational Linguistics*, Oxford University Press, 2003, pp. 599-615

JACQUEMIN, C., ET TZOUKERMANN, E. (1999), NLP for Term Variant Extraction: A Synergy of Morphology, Lexicon and Syntax. In T. Strzalkowski, editor, *Natural Language Information Retrieval,* pages 25-74, Kluwer, Boston, MA, 1999

KAHAN J., KOIVUNEN M.-R., PRUD'HOMMEAUX E., ET SWICK R.R. (2001) *Annotea : an open RDF Infrastructure for Shared Web Annotations*, Proceedings of WWW10, May 1-5 2001, Hong-Kong, pp. 623-632.

KINTSCH W. (1988). The role of knowledge in discourse comprehension: A Construction-Integration model. *Psychological Review*, 95, 163-182.

MARCOCCIA M., (2004) *On-line polylogues: conversation structure and participation framework in internet newsgroups*, Journals of Pragmatics, 36 (2004) 115-145.

PÉDAUQUE, R.T. (2003). Document : forme, signe et médium, les re-formulations du numérique, *working paper*, version 3- 8 juillet 2003, http://rtp-doc.enssib.fr

PIOLAT A., FARIOLI F., & ROUSSEY J.-Y. (1989). La production de texte assistée par ordinateur. In G. Monteil, & M. Fayol (Eds.), *La psychologie scientifique et ses applications* (pp. 177-184). Grenoble : Presses Universitaires de Grenoble

ROUSSELOT, F., FRATH, P., ET OUESLATI, R. (1996), *Extracting concepts and relations from Corpora.* In Proceedings of the Workshop on Corpus-oriented Semantic Analysis, European Conference on Artificial Intelligence, ECAI 96, Budapest, 12 August 1996.

SUMNER T., BUCKINGHAM SHUM S., WRIGHT M., BONNARDEL N. , PIOLAT A. & CHEVALIER A. (2000). Redesigning the peer review process : A developmental theory-in-action. In R. DIENG, A. GIBOIN, G. DE MICHELIS & L. KARSENTY (EDS.), *Designing cooperative systems: The use of theories and models* (pp. 19-34). Amsterdam : I.O.S. Press

TCHOUNIKINE P. (2002-a). Pour une ingénierie des Environnements Informatiques pour l'Apprentissage Humain. *Revue I3 Information-Interaction-Intelligence*. Vol. 2, n°1, Cepadues Editions.

TCHOUNIKINE P., (2002-b), Conception des environnements informatiques d'apprentissage : mieux articuler informatique et sciences humaines et sociales, in Baron G.L., Bruillard E. (eds.), *Les technologies en éducation : Perspectives de recherche et questions vives*, p. 203-210, Paris : INRP - MSH - IUFM de Basse Normandie.

WEICK K.E., (1979) *The Social Psychology of* organizing, New York, Random House.

ZACKLAD M., LEWKOWICZ M., BOUJUT J-F., DARSES F., ET DÉTIENNE F. (2003), *Formes et gestion des annotations numériques collectives en ingénierie collaborative*, actes des journées Ingénierie des Connaissances 2003, Laval.

ZANNOT (2003), http://www.zope.org/Members/Crouton/ZAnnot/

Il ne faut pas numéroter les pages -