Annotation: textual media for cooperation

G.Lortal[1], M.Lewkowicz[1], A.Todirascu-Courtier[2]
[1]*Laboratoire ISTIT équipe Tech-CICO, Université de technologie de Troyes*
[2]*EA 1339- Linguistique, Langues et Parole, Université Marc Bloch - Strasbourg*
[1]*{lortal, lewkowicz}@utt.fr*
[2]*amalia.todirascu@umb.u-strasbourg.fr*

ABSTRACT

This paper aims at describing an annotating model used to underlie a collaborative annotation tool development. Annotating becomes more and more important, as digital documents become central in work situations. Supports for deliberation and arguing around these documents are needed. In this article, we propose a definition of annotation merging Semantic Web view of annotation (tagging) and Social Web view of annotation (comments). We are also presenting existing annotation tools classified according to the Computational/Cognitive dichotomy of annotation definition that we propose, and conclude on the need to develop an annotation tool merging the social annotation to enable discourse about a document in a distributed environment and to support it by way of indexing (computational annotation). Our design process for this annotation tool consists in starting from an annotating model inspired from humanities theories. After a review of main models fitting to our purpose, we conclude by deploring the lack of interaction representation in cognitive models of comprehension or text creation, and the low focus on the annotating production. We then propose a model inspired from the medieval rhetorical discourse production. We finally detail its ongoing implementation and discuss how it fits basic features of cooperative annotating.

KEYWORDS
Annotation, Computer Supported Collaborative Work, Annotating Model, Rhetoric, Hermeneutics

1 ANNOTATING DURING ASYNCHRONOUS WORK PHASES

In a context where mediated exchanges increase, digital documents become central; people need to share written documents and to share ideas by writing. In design activities, digital documents are the basis of deliberations and confrontations. In fact, when working in a cooperative way, interaction among the project's participants is crucial. During asynchronous work phases, this interaction could take place through documents' reviews, the digital document supporting episodes of collective arguing in addition to face to face exchanges (Darses, 2001), (Martin *et al.* 2001). Documents are then always in action, they keep changing, and they bring about changes, such as messages in a conversation. Annotations could then be seen as messages posted around a document, enabling arguing. An annotation is then a fragment of a Document for Action (Zacklad, to be published).

These critical comments through annotations are useful especially in long-term processes as they represent a mnemonic support for knowledge capitalization during a design project. Annotations enable tracing the Design Rationale, decisions' follow-up. Finally, using annotations or glossing comments promotes interpretation by allowing memorizing and recall, and then takes part in collective interpretation, the collective sensemaking (Weick, 1979).

In this paper, we first explain our positioning concerning annotation and annotating. Then, we check existing tools and models reusable for our purpose. And we finally expose our model and describe the tool which is to be developed according to our presented features.

2 ANNOTATION: RESULT OF AN ANNOTATION ACTIVITY

2.1 Annotation types

Annotation can enrich documents on several levels. (Handschuh and Staab, 2003) define shallow and deep annotations on the Web, distinction made upon the one between static and dynamic documents. The last ones are composed of a structure (deep) and of an external content (visible by the reader). A

deep annotation concerns the static part of a dynamic document and a shallow one finds itself in a static document. These authors only take into account Semantic Web (SW) annotation i.e. annotation used for document content processing by the addition of embedded commands. As recommended by Semantic Web W3C (World Wide Web consortium), annotation can be used to search for information, to structure a document, to shape a document, to enable services interoperability, and for certain types of cooperation (as in (Koivunen and Swick, 2003)'s scenarii of cooperation). We claim that this is an annotation definition at a *computational level.* Indeed, annotation can also be seen, as we claimed in introduction, as a type of Document for Action (Zacklad, to be published) (DOfA in the following). The annotation is a "documentarized" fragment of a DOfA which adds supplementary information to a document, contributing to collective sensemaking. Within this definition, we are pointing out a broader sense of annotation. Annotation is not only a SW tool, but also a cognitive means. At the *cognitive level*, annotation is used for information processing by human beings. The cognitive processing of content can take place within a collective context as well as an individual context. Annotation is a clarification, it allows the emergence of a common vision of an object which is still under construction, as well as the emergence and the reinforcement of the identity of each participant inside the group. Annotation adds semantic information, which helps to support a cognitive relation with the DOfA.

We can then define annotation as a metadata, a boundary object, semantically enriching a document. The document is then enriched by the relation, the path between the content of a document and other contents. This other content could be a computationally calculated content (the HTML visualisation **bold** by means of the markup / tag <b>), another document (a hyperlink, a bookmark), a Semantic Web or syntactic markup (XML tags, language parsing) or even the textual body of an annotation ("I disagree with your view of the Semantic Web"). Annotation could be seen as a continuum from the markup to the comment (fig.1).

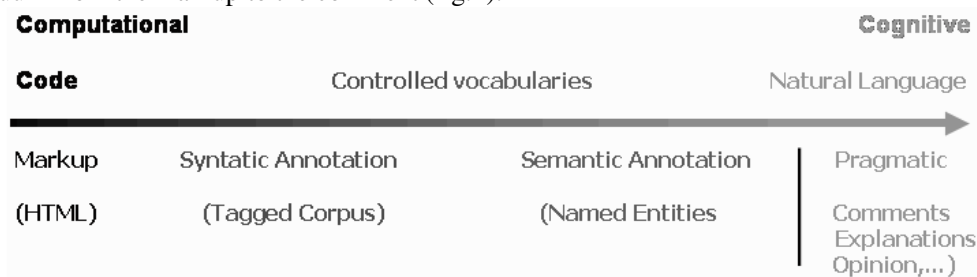| Computational			Cognitive
Code	Controlled vocabularies		Natural Language
Markup	Syntatic Annotation	Semantic Annotation	Pragmatic
(HTML)	(Tagged Corpus)	(Named Entities	Comments Explanations Opinion,...)

Fig.1 Annotation Continuum

We are particularly interested in the cognitive enrichment of a document, defining annotation as a discourse fragment in connection with a text, an argument medium. We will focus on supporting this annotation activity with a tool in order to assist collective sensemaking.

2.2 Defining annotating activity

Annotation is a traditional element of hermeneutics (De Libera, 2000) and of rhetorical discourse. A discourse in rhetoric is a chain of arguments. Indeed, we are considering the annotation definition in a cognitive view, as a textual fragment anchored to the document, which arouses an assessing idea. In this scope, annotation allows distant actors to propose changes in a document and to interact with others. Annotation is an activity as well as the final product of this activity. We will now call this activity *annotating.* Annotating is a writing activity arising out of reading, in other words it is a reading-writing activity which enables arguments' exchanges within a group.

As a cooperative activity, annotating represents the shaping of ideas, materialized by text fragments which allow communication around a document. The production of a textual fragment could be seen as an individual process, but the text produced by appended fragments encloses participants' arguments about a document. Appended fragments are enclosing a context to the text, contextualising a text by including some production and reception conditions among the text (Charaudeau and Maingueneau, 2002). That's why we propose to consider this text production as a discourse production following the traditional linguistic dichotomy between text and discourse. This discourse production is a collective cyclical process since the read document is developed or even

rewritten before being given to a group. The resulting product of this activity is a discourse fragment, a discursive annotation, anchored to a/several document(s). The fragment is then bound to the document(s), but also to other fragments bound to the same document or to the fragment itself.

Indeed, to support this activity, we need to use relevant groupware supporting a critical reading-writing activity, also called hermeneutics. This tool should mediate discourse production through annotations in a collective activity, in other words it should enable a hermeneutical way of annotating. We are now going to expose existing tools focusing on this problematics.

3   EXISTING TOOLS SUPPORTING COGNITIVE ANNOTATION

We are presenting here some of the most representative annotation tools previously described in (Lortal *et al*, 2005a). Nowadays, several annotation clients are available, stemming from SW initiatives.

Most of them adopt the term "semantic annotation", that we would call "computational annotation". The main objective of this approach is to index web pages more or less automatically. In fact, metadata are added in order to index a web page, and allow search engines a better information or pages recall. These tools are used for metadata creation and some are based on ontologies to support computational annotation, for example OntoMat-annotizer (Handschuh *et al.*, 2002), Melita (Dingli, 2003), MnM (Domingue *et al.*, 2002). Computational annotations only enrich a web page with concepts for automatic indexing and do not either allow readers of a same page to cooperate or to interact. These computational annotations could also be used for another use purpose, as in KMI's Magpie (Dzbor *et al.*, 2004), which uses computational annotations to support human interpretation of web pages.

In our view, annotations are not only computational but also discursive. Thus, our purpose is more to support the creation of new ideas (from collective interpretation) than to strictly support document interpretation or recall of an existing interpretation. Then, we cannot content ourselves with computational annotations, even if shared by others, because they only help user structuring her/his mind and understanding, or share her/his understanding of the text. We need to find tools managing *discursive* annotations that is to say enabling discourse around a document.

Another type of annotation clients adopting a more social approach, could be interesting. They aim at facilitating human communication, without considering indexing features or annotation recall. In these clients, annotations can only be sorted on rudimentary metadata such as the creation date or the author, as for example in Yawas (Denoue, 2000), CritLink (Ka-Ping, 1998), XLibris (Price *et al.*, 1998). These tools assume that the annotation is a comment, a way of looking at annotations shared by some proprietary software or some plug-in application software, where the comments are neither indexed nor differentiated from the document (Windows Word comments, 2003). Annotations are sometimes stored apart on annotation servers (Acrobat pdf, 2004) and organized in a minimalist way. However, these tools do not allow connecting annotations, and they cannot represent a structured set of exchanges between users related to a document.

We are sharing the KMI's D3E (Sumner *et al.*, 2000) view, considering documents as discourse medium. However, this tool does not allow a rich indexation of annotations, and then it will be difficult to the participants to understand the Design Rationale underlying the discussion. Thus, even if D3E supports interaction better than computational annotation tools, it is not sufficient for our purpose.

To sum up, we can classify tools supporting annotating in two families; one concentrates on the Web pages indexation, supporting their recall, while the other one concentrates on the human communication through comments. We can deplore the lack of annotations' management or the poverty of cooperative functions in these two families, even if KMI's propositions are the first steps in linking these two points of view. Following them, we thus propose to design a tool mixing functions supporting annotating activity in a cognitive way (answering an annotation, or multi-anchoring of annotations for example, as we will explain below), and SW indexing techniques.

Moreover, another weakness of annotation tools is that, in fact, they are a new category of systems handling a new type of mediated activities. As this field is emergent, tools developed for annotation purposes are not relying on a modelling thought about the activity being assisted, because of the lack of existing observable practices. In order to fill this lack, we propose a design process based on an underlying model inspired from humanity theories. As we already said above, we claim

that annotating is a hermeneutical activity, which means that the underlying model representing annotating activity has to portray a critical reading-writing activity. In the following part, we are going to expose candidate models.

4   EXISTING MODELS ENABLING STRUCTURING AN ANNOTATION TOOL

We propose a tool supporting user's annotating activity based on a theoretical model describing this activity. This model should describe a collaborative reading-writing process, which will be implemented step by step in our annotating system. As annotation is seen as an object coming from comprehension as well as production process, we are going to present some reading and some writing models mainly stemming from Cognitive Sciences. First of all, we looked at the (Kintsch & Van Dijk, 1978)'s comprehension process model, then, (Hayes & Flower, 1980) and (VanWijk & Sanders, 1999) production process models. As these models do not fit collaborative aims, we have looked at the "social interaction model" from (Nystrand, 1989) fitting collaborative purposes and representing the whole interaction between a writer and her/his readers around a text. Nevertheless, this last model cannot explain the annotating process taking place in the whole text production model. Then, we finally present a rhetorical model which we adopt.

4.1   Cognitive models

An exhaustive list of cognitive models describing reading and/or writing activities is not our purpose here, we will rather focus on models widely used in CSCW, or supporting collaborative annotating. Reading or writing models stemming from cognitive sciences, and mainly from cognitive psychology, aim at functionally represent a given process, i.e. reading or writing. So they finely describe process functions, elements or relations.

Reading is often seen as a comprehension or a memorization process. But since we consider a mediated activity focusing on DOfA (for example a collaborative design activity), it seems difficult to completely dissociate this comprehension process from a writing one. In fact, documents are collaboratively written or negociated, that is to say that several people are writing in a single document device. This cooperative process is marked out with written traces. This acknowledgment leads us to the necessity of using cognitive models focused on expression *and* written processes. Indeed, we assume that our model should represent one unique process of comprehension-expression. In the case of a design team, writing while designing could be seen as acting throughout a comprehension or an interaction within a group. Annotating could also be seen as a comprehension process trace along a reading. Annotation aims at sharing comprehension within a group. Moreover, traces as a whole enable the group to form a temporarily shared social reality (TSSR, (Rommetveit, 1974)).

The (Kintsch and Van Dijk, 1978) comprehension model is well-known and used in cognitive psychology, and number of experiments reinforced the propositional theory of Kintsch and Van Dijk. These authors proposed a semantic analysis of narratives funded on clauses (sentences). A clause is made up of a predicate (an acting relation) and one or several arguments. This analysis of a story in terms of clauses identifies micro-structures and macro-structures. Kintsch and Van Dijk propose macro-rules (generalization, deletion, integration, construction) to go from a story expressed in clauses to micro-structures, and then a macro-structure. Most experiments concerning this model are related to the activity of producing stories' summaries. These experiments emphasize clauses' transformations and clauses' re-organization processes in a text which has already been presented, and measure information recall rates. Even if principles underlying this model seem interesting, experiments still stress on the recall quality and quantity, and not on text creation, or on the building of new ideas or concepts.

In our cooperative frame, the activity concerns more new predicates creation–i.e. expression- than the information recall which has already been presented, bound to a narrative production schema –i.e. comprehension–. Then we are going to present some general production models, and some writing production models.

One of the main references on this topic is the Hayes and Flower model (Hayes and Flower, 1980; Hayes et al., 1987; Hayes, 1995). It is a founding editorial model widely acknowledged by the cognitive science community (Piolat, 2004). This model is based on three modules: (1) the task environment, (2) the long-term memory, (3) the one of strictly speaking writing process. The writing

process consists of three tasks getting on recursively. The first task, the planning phase, consists in organizing writer's ideas according to its objectives. Then, the translation represents the preverbal message (formed during the planning phase) encoding. Encoding the message means applying a code which could be at a graphemic, spelling, lexical and syntactic level. It produces the message read or listened to, by peers. Thirdly, the review process takes place. It is a control process defined in (Bereiter and Scardamalia, 1987) as inadequacy detection with regards to writing constraints or foreseen text mental representation, inadequacy's characterization, and inadequacy's correcting strategy choice. Interaction between cognition and environment is underlined in (Hayes, 1995) quoted in (Barré de Miniac, 2000) but the model remains individual. The activity result (in our case, annotations) is only one of the elements of the environment. In a mediated context, annotating is an activity firmly stamped with its co-text (textual context) and context, especially with the cooperative context generating interactions between the group/project members.

As annotation is a contextual object, we wonder if cognitive model stemming from oral production could better fit our purpose than strictly written production model. People often see that oral discourse is more context-dependent than writing, then we can assume that general production model would rely more on environment.

The (Van Wijk and Sanders, 1999)'s model propose a writing model based on (Levelt, 1989) oral verbal production model. This model is proposing two interesting alternatives. First of all, Van Wijk and Sanders, inspired by Levelt's works in psycholinguistics, underline similarities between written and oral production. Several aspects of verbal production (written and oral) are known as common. From a theoretical perspective, this model could be applied to what is considered as a new language style, the mediated language - a blend of oral style and written style (Herring, 1999, Muniandy, 2002). Secondly, this model explicitly merges the comprehension and expression phases into a single process.

The weakness of all these models remains the lack of collaborative processes during the production process. They stay focused on a cognitive process, an individual and internal production process.

4.2   A social interactive model

Another way of studying writing processes is to focus on the social characteristic of writing. According to (Nystrand, 1989) writing nature shares with language nature its social property. Criticizing (Hayes and Flower, 1980) for the poverty of the interacting processes during the translating phase, Nystrand stresses the interface between cognition and text. The audience, which was only an environmental constraint in (Hayes and Flower, 1980), becomes a central element in the social interaction model.

The social interaction model presented in (Nystrand, 1989) is rooted in oral communication. Nystrand sets his model in an interacting context built by a writer and her/his audience (to which the writer belongs her/himself). In this sense, interaction occurs each time a reader understands a written text. We are then close to the hermeneutical situation proposed in (Gadamer, 1996). In this communicational view of the writing process, Nystrand considers the written production as a sense negotiation between a writer and its reader, in order to create a common reference framework. This model proposes a three-phase editorial activity representation. The three operations are mainly carried out by the writer: (1) discourse initialization, (2) discourse maintenance/readjustment, and (3) elaboration. This model is interesting regarding exchanges' relations between a writer and a reader. In fact, the interaction between peers is considered as a strong requirement and the exchange process is central to the writing process. Nystrand also states that text is not only the result of a composition, but is also a communication medium.

However, Nystrand's model is a perfect triangle relation among writer, reader and text; communication is seen only chronologically and as forming a whole. In an annotating situation, we cannot find this type of production. Annotation is not a whole structured production from a writer to its reader, but is fundamentally bound to its context, and not only its co-text. Annotation can be seen more as an answer from the reader to a writer, the swapping of these two roles. It is a readjustment element of the discourse. In (Nystrand, 1989)'s model, we miss the dynamic facet of the annotation, the fragment of discourse giving rise to action (the DOfA, (Zacklad, to be published)).

4.3 A discourse production model

The first models about discourse production can be found in Aristotle's rhetoric description. A discourse is, at the outset, written, following some well-speaking rules, then it is learnt to be publicly declaimed. Aristotle's model has been improved to fit with a new kind of rhetoric fulfilling hermeneutic objectives i.e. text interpretation by discursive exchanges during some public academic events. This public arguing exercise is strongly addressed and represents a blend between oral and writing as well as a blend between comprehension and expression. In fact, to explain this model, a reading phase has been added, describing the importance of making links between ideas when reading, and ideas already in memory, and the importance of organizing not only the text, but also ideas in memory.

Taking into account "the others" is not only done during the expression phase; the whole discourse is defined in an interactive way. The first phase (reading-comprehension) of the model is also interactive because knowledge comes not only from the read text, but also from people listened to. Knowledge is, in any case, generated by relation with other text(s) or idea(s), or by interaction with other people. The main focus of the model is this perpetual adjustment with context, a kind of hermeneutical circle by means of regenerating relations according to what is in memory and what is perpetually accessed. We are dealing with a continuous text production, stamped with context interaction.
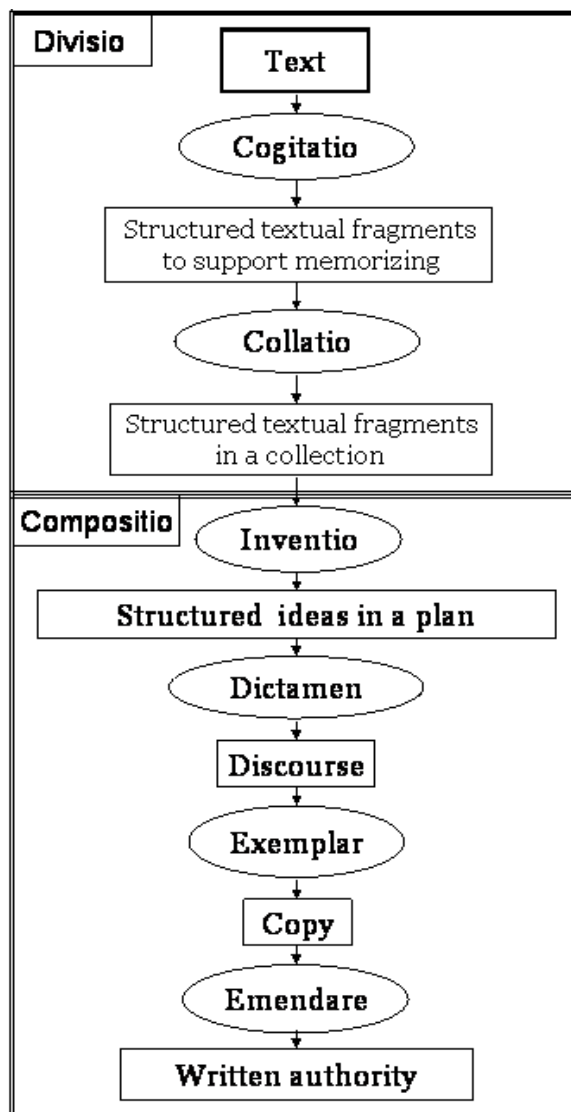

Fig. 2 – Discourse production model

The discourse production process as recommended in this context of medieval rhetoric is made up of two phases: "Divisio" and "Compositio". Divisio is done while reading, and consists in dividing a text into understandable units, in memorizable short segments. Compositio is the ordered combination, the suitable arrangement of "res" (conceptual or material objects) contained in the memorized segments (Fig.2). These memorizing - Divisio - and creation – Compositio - phases, are themselves divided into stages.

The first stage of Divisio is Cogitatio. It is an individual memorial stage which consists in associating (by a conscious choice and recall) images and sections of a chronologically divided content of a document in various memory locations. Textual fragments that form the text are then structured and become easily memorizable. Collatio is the phase where textual fragments stored in several distinct places in memory are combined in a structure. In this phase, connections between the various fragments are created. A co-text is then formed by semantically binding new memorized fragments and fragments previously memorized.

In this rhetorical model, Cogitatio and Collatio are supported by individual/mental "notae" indexing fragments. This Divisio phase is seen as individual because it structures an individual memory, but in an asynchronous way of working, this phase

should be mediated in order to allow discursive exchanges i.e. interactions with others.

Compositio is divided into four stages of activity, evoking stages of textual document creation. The stage of Inventio is close to that of Collatio insofar as it focuses on creating semantic links between various memorized elements, on the "res" (conceptual objects, idea) level, and not on the word level. An outline is formed, i.e. a set of ideas hierarchically organised (for example, an argument structure).

The following phase will be the word-level formatting of this conceptual outline. It is a traditional phase of drafting, called "Dictamen". We see with this stage the physical discourse creation, classically done on an adjustable support (a draft), where the style, the choice of the terms, therefore only the textual shape of the discourse can be modified.

The Exemplar phase consists in transforming the draft support of the discourse in a perennial support. The discourse remains strictly identical to the one found in output of the process of Dictamen.

The last phase in this process is Emendare, where the final copy of the discourse is diffused and then openly commented by the addition of public comments, arguments or annotations of an author to the original text. After this phase, the text becomes a reference text, a written document being an authority on the field.

This model stemming from medieval rhetoric describes the production of a document, traditionally textual and individual, so synchronous. In an asynchronous or distributed activity of writing, the Divisio stage during which the text is read and the author is structuring his/her ideas, can not follow the traditional choices of rhetoric. In so doing, we would let a user alone in her/his understanding, that is to say letting her/him making weak individual relations between ideas, and building fewer ideas than if there were several users exchanging explicit and explained relations. These relations could be realised through annotations along the body of the text and throughout the text comprehension.

5   WHEN NOTAE BECOME ANNOTATIONS

In the discourse production model that we propose, we can spot notae (or annotations, when mediated) phases. We propose to see the discourse production model presented in figure 2 as the representation of the global document (DOfA) development, the annotation phase being mostly represented by the Divisio phase, where Cogitatio and Collatio occur. These phases are typically the ones where readers/authors exchange comments about a document, and mark "notae" indexing and arranging their ideas and concepts. Traditionally, fragments arise out of the reading of the text, and interactions between author's ideas and her/his readings (Cogitatio) and authors and her/his fellows (Collatio). Readers/authors are "committing" changes on the read document. They contextualize it by creating a context to the document through notae in adding fragments. Adding a nota means modifying the context, and so the document. Nota could be seen as a kind of indexing object, a "metadata" or as a contextualizing object, a "co-text" and a context.

In so doing, readers/authors "commit" a new version of the DOfA. In an asynchronous and distributed context of work, we should be able to track down fragments' elaboration marked off by notae and so be able to store clues of collaboration, i.e. notae that enabled discourse production as well as collective sensemaking. These notae are typically what we called annotations, considered as marks of versioning on a DOfA, enabling alteration, amendment of a document in a collective work.

In a mediated discourse production activity, the annotation object is the product of a mediated annotating activity as described above. The object is a textual fragment arising out of phases aiming at structuring segments (text segmentation, Cogitatio) or structuring concepts (text indexation, Collatio).

For example, in a cooperative work context, one can consider the sharing of a document in order to be commented on. After a visualization phase of the text (a reading) the read will be segmented to allow the addition of a structured comment, of a discursive annotation. A segment will be emphasized in order to indicate the anchoring of a discursive element linked to this segment. This highlighting could be done by traditional techniques of underlining, circling, colouring segments of unsettled sizes (from a word, or a part of a word, to the paragraph, or set of separated elements). This phase mainly needs visualisation techniques.

Following the segmentation and the choice of element to be annotated, an indexing phase is required, consisting in connecting segments. The tool should help the user to find semantic links between elements to structure them together and to form an organized set of textual segments

according to their meaning. This meaning depends on the user's understanding. Indeed, the annotation consists in an anchor, a geographical relation, in a body, a discourse which creates its meaning amid a "co-text", but also in the whole set of textual segments stored in memory and linked to it, indexed to it by comprehensible keywords, structured by and for human users. All these indexing problems are tightly bound to browsing.

While writing the annotation, the author should also organize his/her ideas to be written. This necessary step is the structuring of "rei", concepts stored in memory, which will give rise to an outline made up of hierarchically structured arguments. The writing phase will produce the body of the annotation which will be readable by a member of the discussion after publishing and thus spreading this annotation. Just as a reference text, the annotation can be endorsed thanks to a new link brought to the latter. A reply to a comment allows taking part in the discussion thread initiated by the first annotation. Annotation is in many ways a rewriting of a document fragment. Its objective is to broaden, to explain, to bind a/several document(s) fragments. It is a matter of interpretation.

This text fragment is central to the discourse production activity. As a single interpretation element, annotation is anchored to a document fragment. But as one of several interpretation elements, it is bound to one or several document fragments and/or other interpretation fragments. In a cooperative view, annotation of an author answering to annotation of another author in a set of documents is more a discourse fragment. The discourse is then composed of the set of annotations bound to the same document, the same theme, the same author, etc. This homogeneous set of annotations, this new discourse about a document, is the interpretation developed and it "commits" the document as a Document for Action (DOfA). The new document created by gathering of homogeneous annotations is itself a DOfA which can be commented again as a document produced following the discourse production activity model. Reaching this point, we are touching the second important phase of cooperation through annotation, the rhetoric phase of Emendare aiming at public comment of a document. We will not deal with this point here but we can underline the functionalities that the tool should implement.

6 REQUIREMENTS AND TOOL

In order to assist users in annotation creation, we adapt a model of medieval rhetoric (Carruthers, 1990), representing discourse creation, to our problematics of textual fragment creation, the activity of annotation. The user, by anchoring textual objects to a document, carries out a specific activity that we describe through this rhetoric model at several stages. On the basis of this discourse production model, we define a mediatized discourse production model where the annotation activity is mediated via the use of a tool (Lortal *et al*, 2005b). From this model, we release some design primitives for a tool supporting collaborative text interpretation. These primitives are essential functionalities to support users in its annotation creation as well as in annotation visualization aiming at a possible information re-use. These primitives are basic functionalities of a tool implementing indexing solutions based on Natural Language Processing (N.L.P.) techniques for a cognitive classification.

Having described these activities (section 5), we can draw actions to be supported by a groupware to allow users annotating in cooperative activities.

We can consider three main functionality families: interpretation, browsing, creation. Supporting interpretation means handling annotations as creating fragments of discourse and enabling discourse by creating threads as answer of an annotation for example. Functions as selection of document fragments (highlighting, circling…), anchoring discourse fragments documents and fragments (answering, multi-anchoring…), are then necessary.

Once created, this annotation to be recovered and structured should be indexed. Browsing is based on annotation indexing. Indexation allows structuring annotations in browsable knowledge map as Topic Maps formalism allows (Biezunski *et al*, 1999). To index subtly these fragments, the user should be involved. But to support the user in this time-expensive task, we suggest using N.L.P. tools proposing user domain specific terms and the annotation arguing type.

Thirdly, users should be able to create new documents which gather ideas emerging from collective brainstorming and exchanges around a document. Our tool should contain a gathering functionality allowing creation of a draft to work on.

7 CONCLUSION

Within the aim of an iterative groupware design, a version of the tool is under development with Open Source technology. It is based on the W3C's annotation standard (Annotea, 2003) and the annotation server (Zannot, 2003). It contains all functionalities defined above, and it respects a distributed architecture enabling collective work, and follows W3C standard and ISO norms.

It will allow an evaluation of our assumptions on the discourse production model and on the annotation status. In order to validate an annotation typology, an experiment is now being carried out involving a group of mechanics researchers collaborating in synchronous and asynchronous distributed phases, through plans for the design of a plane engine. We study the exchanges carried out while asynchronous phases of work in order to trace design rationale of the activity via annotations and to create a corpus on aeronautics. This corpus will be used to train N.L.P. tools used for indexation purposes.

8 AKNOWLEDGEMENT

9 REFERENCES

Acrobat PDF, (2004) http://www.adobe.com/support/techdocs/ac76.htm

Annotea, http://www.w3.org/2001/Annotea/ (2001)

Aristote, *La rhétorique*, Livre 3, livre de poche, 1991

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.

Bereiter, C. & Scardamalia, M. (1987). *The Psychology of Written Composition*. Hillsdale, NJ:Erlbaum

Biezunski, M., Bryan, M., et Newcomb, S. R., (1999) *« Topic Maps », spécification ISO/IEC* 13250, 3 Décembre 1999.

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

Charaudeau P. and Maingueneau D., (2002) article Discours in *Dictionnaire d'analyse du discours*, Seuil.

Darses, F. (2001) Converger vers une solution en situation coopérative de conception : analyse cognitive du processus d'argumentation. In F. Darses (Ed.) *Modeling Cooperative Activities in Design Proceedings of the 10th Atelier du Travail Humain*, 27-28 juin 2001, Paris, France: INRIA.

De Libera A., (2000), *La philosophie médiévale*, Paris, PUF (« Que sais-je ? » 1044), 4e éd.

Denoue, L., et Vignollet, L. (2000), An annotation tool for Web browsers and its applications to information retrieval, in *proceedings of RIAO 2000*.

Dingli A., (2003), Next Generation Annotation Interfaces for Adaptive Information Extraction. In *6th Annual Computer Linguists UK Colloquium (CLUK 03),* January, 2003, Edinburgh, UK, 2003.

Domingue J.B., Lanzoni M., Motta E., Vargas-Vera M., et Ciravegna F. (2002), Mnm: Ontology driven semi-automatic or automatic support for semantic markup. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02),* October 2002.

Dzbor, M. - Motta, E. - Domingue, J. B. (2004), Opening Up Magpie via Semantic Services. In *3rd ISWC*, November 2004, Japan

Gadamer H-G. (1996) *Vérité et méthode. Les grandes lignes d'une herméneutique philosophique* (1960) ; trad. Fruchon, Grondin, Merlo, Seuil, 1996 ; Vol.1 des Gesammelt Werke, Mohr, Tübingen, 1986.

Handschuh S. and Staab S. (2003) Annotating of the Shallow and the Deep Web, in *Annotation for the semantic web*, S. Handschuh, S. Staab, IOS Press p.25-45

Handschuh, S., Staab S., et Ciravegna, F. (2002), S-cream - semi-automatic creation of metadata. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, October 2002.

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.

Hayes JR (1995). Un nouveau modèle du processus d'écriture. In J. Boyer, J-P Dionne et P. Raymond (dir), *La production des textes. Vers un modèle d'enseignement de l'écriture*. Montréal : Les éditions logiques, pp 49-72.

Hayes, J.R., Flower, L., Schriver, K.A., Stratman, J.F. & Carrey, L. (1987). Cognitive processes in revision. In Rosenberg (Ed.), *Advances in applied psycholinguistics*, (vol.2, pp. 176-241).

Herring, S. (1999). *Coherence in CMC. Journal of Computer-Mediated Communication*, 4(4). Retrieved January 12, 2005, from http://jcmc.indiana.edu/vol4/issue4/herring.html

Ka-Ping Y. (1998), *CritLink : Better hyperlinks for the WWW*. http://crit.org/ping/ht98.html, 1998.

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

Kintsch, W. & Van Dijk, T.A. (1978). Toward a model of text comprehension and production, *Psychological Review*, 85, 363-394.

Koivunen, M-R, and Swick R.R. (2003) Collaboration through annotations in the Semantic Web in *Annotation for the semantic web*, S. Handschuh, S. Staab, IOS Press p 46-60.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA : MIT Press.

Lortal G., Lewkowicz M., Todirascu-Courtier A. (2005a). AnT&CoW, a tool supporting collective interpretation of documents through annotation and indexation, in *Proceedings of KMOM-IJCAI Workshop* p.43-54.

Lortal G., Lewkowicz M., Todirascu-Courtier A. (2005b). Modélisation de l'activité d'annotation discursive pour la conception d'un collecticiel support à l'herméneutique, in *Proceedings de la IC2005 conference* p. 169-180.

Martin, G., Détienne, F., Lavigne, E. (2001) Analysing viewpoints in design through the argumentation process. In *proceedings of INTERACT'2001*, Tokyo, Japan, July 9-13.

Muniandy, A.V.A. (2002). Electronic-discourse (E-discourse): Spoken, written or a new hybrid? *Prospect: An Australian Journal of TESOL*, 17, 45-68.

Nystrand, M. (1986). *The structure of written communication: studies in reciprocity between writers and readers*. Toronto: Academic Press.

Nystrand, M. (1989). A social-interactive model of writing. in *Written Communication*. 6, 65-85.

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

Piolat, A. (2004). Approche cognitive de l'activité rédactionnelle et de son acquisition. Le rôle de la mémoire de travail. *LINX (Linguistique Institut Nanterre Paris X)*, 51, 55-74.

Price, M., Schilit, B., et Golovchinsky, G. (1998), XLibris: The active reading machine. In *proceedings of CHI'98 Human factors in computing systems*, Los Angeles, California, USA, vol.2 of Demonstrations: Dynamic Documents, pages 22-23, 1998.

Rommetveit, R. (1974). *On message structure: A framework for the study of language and communication.* London: John Wiley.

Sumner T., Buckingham Shum S., Wright M., Bonnardel N., Piolat A. & Chevalier A. (2000), Re-designing 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, 2000.

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

Wijk, C. van, & Sanders, T.J.M. (1999). Identifying writing strategies through text analysis. *Written communication : a quarterly journal of research, theory and application*, 16, 51-75.

Windows Word (2003), http://office.microsoft.com/fr-fr/assistance/HA010714941036.aspx

Zacklad, M. (to be published) Documentarization processes in Documents for Action (DofA): the status of annotations and associated cooperation technologies, in *JCSCW…*

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