

Including Authorial Stance in the Indexing of Scientific Documents

Viviane Clavier* and Céline Paganelli**

* Université Stendhal – Grenoble 3, Institut de la Communication et des Médias, 11 av. du 8, mai 1945, BP 337, 38434 Échirolles, <Viviane.Clavier@u-grenoble3.fr>

** Université Paul Valéry, ITIC- Département Information Documentation, Route de Mende, 34 199 Montpellier cedex 5, <Celine.Paganelli@univ-montp3.fr>



Viviane Clavier is Associate Professor in Information Science & Communication Studies at Grenoble3 University. She is a member of Gresec (research group in Information and Communication Studies, Grenoble, Stendhal University). Her research interests revolve mainly around information retrieval and indexing of scientific and professional discourse.



Céline Paganelli is Associate Professor in Information & Communication Studies at Montpellier 3 University. She is a member of Gresec (research group in Information and Communication Studies, Grenoble, Stendhal University). Her research interests revolve mainly around information uses in professional context. She has studied scientists, bankers, and researchers' information practices, in particular.

Clavier, Viviane, Paganelli, Céline. **Including Authorial Stance in the Indexing of Scientific Documents.** *Knowledge Organization*. 39(4), 292-299. 41 references.

ABSTRACT: This article argues that authorial stance should be taken into account in the indexing of scientific documents. Authorial stance has been widely studied in linguistics and is a typical feature of scientific writing that reveals the uniqueness of each author's perspective, their scientific contribution, and their thinking. We argue that authorial stance guides the reading of scientific documents and that it can be used to characterize the knowledge contained in such documents. Our research has previously shown that people reading dissertations are interested both in a topic and in a document's authorial

stance. Now, we would like to propose a two-tiered indexing system. Dissertations would first be divided into paragraphs; then, each information unit would be defined by topic and by the markers of *authorial stance* present in the document.

Received 7 March 2012; Accepted 7 March 2012

1.0 Introduction

This article will focus on the indexing of scientific documents in connection with the context of their use. Working in the field of specialized information, we believe that knowledge is the product of an encounter between the data contained in a document, a user in search of information, and a specific context. Context may refer to an organizational, technical, or human environment. Knowledge is, as such, understood as both the result and interpretation of data in connection with users, their activity, and a given con-

text. As such, issues surrounding data collection and the representation of knowledge must take into account the nature of a document, the user, and the user's main activity, which we will refer to here as the "context of use."

The research presented in this article follows from exploratory research conducted on doctoral theses readers, the outcome of which was published in the *Les Enjeux de l'Information et de la Communication* electronic journal (Clavier and Paganelli 2010). Our goal was to assess the relevance of the notion of stance in reading and annotating for indexing and

knowledge representation purposes. As an extension of our previous research, we would now like to more specifically point out how linguistic and cognitive knowledge, in connection with stance, can be used to improve access to information.

Stance is not a linguistic category *per se*, but the term is used to designate a series of linguistic processes typical of scientific writing. The large amount of research conducted in the context of research projects is proof of a strong interest in scientific discourse. For example, the Norwegian KIAP project (Kulturell Identitet i Akademisk Prosa)¹ and the French Scientext project focused on scientific writing.² In the latter project, specifically, stance refers to the linguistic processes that reveal “an author’s singularity, their specific contribution – the justification behind their scientific approach – and the author’s reasoning, that upon which the research is based, the proof used, the logical relationships it establishes – the quality of the scientific analysis.”³ We believe that an author’s stance is a driving notion that guides the consultation of scientific documents and is also central to describing their content; as such, we feel that stance needs be a full-fledged part of the indexing process for doctoral theses.

We shall begin with a presentation of the theoretical footing on which our approach is based. Then we will show how the notion of stance is mobilized by users when consulting scientific documents. Finally, we will formulate a certain number of proposals for indexing and the representation of knowledge conveyed in scientific discourse.

2.0 Theoretical Framework

Our approach is part of a body of research from the information and communication sciences.⁴ Given this disciplinary rooting, we have not addressed the representation of knowledge in terms of the formalization of data in the information technology sense of the term; it is not understood as the development of organizational systems in the knowledge management sense either, but it does rely on the description of methods which allow us to draw out data that fuel systems of knowledge representation. There are two trends in the information sciences which differ in how they understand information: recorded knowledge and communicated knowledge. Hubert Fondin has argued that information is part of a process of exchange and sharing, of finalized communication, in a specific context or social system (Fondin 2001, Fondin 2005), and information is, as such, understood as communi-

cated knowledge. Conversely, Yves-François Le Coadic has posited that “information is knowledge recorded in written, oral or audio-visual form on a spatial-temporal medium” (Le Coadic 1994, 6, translated here). For Le Coadic, information is thus understood as recorded knowledge. Our research tends to identify with the first approach since we believe that knowledge exists when there is interpretation, assimilation by an individual and when it is connected to a universe of defined knowledge. Further, we believe that knowledge is constructed by individuals according to the context of use.

This so-called context of use is thus fundamental, both theoretically and methodologically speaking. A lot of research over the past ten years has shown that context has a strong influence on information activity. Brigitte Guyot (2002) has notably shown how information activity is becoming increasingly important in professional contexts. Factors from all levels are involved and influence informational activity— affective states (Kuhlthau 2004) or the specific constraints of a task (Järvelin and Ingwersen 2004)—and a lot of research has focused on information habits in specific professional contexts (Cheuk 1999, Miranda and Tarapanoff 2007, Staii et al. 2006, to name but a few), thus considering that an information activity is affected by context and the activity underway (Bartlett and Toms 2005, Li and Belkin 2008).

This approach has consequences for the methodology behind data collection. We believe that, in some respects, context needs to be taken into account when defining how documents should be processed. This perspective places us within the actor-oriented paradigm (Polity 2000, Chaudiron and Ihadjadene 2002) which includes research that sees information as an interpretive process and that underscores the importance of the concept of context in informational activities (see notably Fidel and Pejtersen 2004, Byström 2007).

We believe that context of use is defined by three variables that have been widely addressed by research, either independently or in a combined manner, and under many different albeit sometimes similar designations, such as the notion of “task,” for example, commonly found in English language research in Library and Information Science (Järvelin and Ingwersen 2004, Byström 2007, Huvila 2008):

- Cognitive factors related to individuals in the context of their work (individual factors: expertise, know-how, the universe of knowledge, etc.);
- Factors related to a person’s professional activity (main activity for which a user is conducting an in-

formational activity, consults or is looking for information in documents; an activity that occurs within a socio-organizational context);

- Factors related to an application (systems, sources of information, documentary genres, specialty fields, etc.).

It is a combination of these three factors that allows us to gather information to represent knowledge. We chose to focus on three sources for the collection, identification, and interpretation of knowledge in order to represent it: documents, users, and the motivations that push an individual to look for information and consult documents. This methodological stage required us to collect “traces,” a term we used to designate data collection methods that allow a corpus to be compiled. Our corpus was defined according to the three sources mentioned above and drew on:

- Documents consulted by users in a professional context, if possible at their place of work. As Dominique Cotte has noted, a document is a very specific object since it is not “data” but rather a “constructed product” resulting from the combination of “signs, alphabets, images, diagrams, [that] can form texts, supported by documents, which may or may not contain information” (Cotte 2004, 31-32, translated here).
- Traces of use or more broadly the “traces of activity” found on such documents (Flon et al. 2009), such as annotations left by a reader on a consulted document or all of the “sources of marking” automatically collected and “redocumented” (Yahiaoui et al. 2011) to explain the “human and social context of activities.” There are various methods for collecting such traces: automatic collection recorded following a computerized action; semi-structured interviews that aim to clarify motivations, the reasons behind the choice of one document, or part of a document over another; and collecting verbal protocols that aim to make subjects “speak out loud” when consulting a document, for example.

This approach was implemented in different contexts, all of which involved a professional situation with users who needed to accomplish a main activity (computer maintenance, writing a thesis, etc.) for which they conducted an information activity. Our previous research conducted in professional contexts (Paganelli and Mounier 2002, Clavier and Paganelli 2010) has shown that information activity is secondary and subordinate to one or more main tasks (preparing a

course, doing computer maintenance, etc.). This leads to different types of reading which are driven by the reader’s goals. Regardless of these goals, reading in a professional context is generally fragmented and non-sequential and involves a large amount of physical and cognitive activity (copy-pasting, underlining, annotations) that leave numerous traces of an individual’s informational activity (Hochon and Jacobini 1994, Mille 2005). In work contexts and depending on the sector, the documents we examined were maintenance manuals, legal texts, medical reports, theses and research articles. Different studies have shown that such documents contain formal characteristics (linguistic and structural) that can be used to improve automatic processing in order to represent the knowledge contained in a document (Péry-Woodley and Scott 2006, Poudat et al. 2006, Couto and Minel 2007).

3.0 Stance as a common thread in the consultation of theses

The way theses are consulted changed a lot when they became available online. The consultation of such documents remains marginal on paper, but has greatly increased for digital versions.⁵ Since 2000, a number of projects and efforts to disseminate electronic versions of theses have emerged,⁶ and such initiatives beg us to think about access methods and the principles of indexing. The question is not new in and of itself. Sylvie Lainé-Cruzel has defined an information system piloted by user profiles for consulting scientific documents (Lainé-Cruzel 1999), and other research has focused on access to French theses in digital libraries (Abascal-Mena and Rumpfer 2007). In the first case, however, access to sources is filtered by the profiles, which is fairly restrictive; and, in the second case, the focus is placed on the semantic content of the document via the extraction of concepts, which limits access to the document’s terminological dimension.

The experiment we conducted has been described in Clavier and Paganelli (2010); it was conducted in three parts. The first phase involved observing the thesis reading habits of ten doctoral candidates in information and communication sciences. Then we questioned them about the criteria they used when selecting theses, and we gathered their comments about the passages of text considered important. We then created a corpus of textual fragments (the passages read) to which we added written annotations from the different media (the actual theses, files, post-it notes, etc.). We also collected oral comments from readers regarding either their consultation strategy or the pas-

sages of text selected. These data were then entirely transcribed and comprised the corpus to analyze.

Among the observed results, it appeared that the consultation of theses by doctoral candidates occurs in a professional setting, in the context of their own research. This type of use corroborates what has been observed in other professional environments (Paganelli and Mounier 2009; Staii et al. 2006): a non-contiguous, often partial reading that leads to an infinite number of experiences influenced by the specific tasks at hand (seeking a definition, problematization, etc.). We observed that approaches to reading differed depending on the number of years a candidate had been preparing their thesis: while readers first seek to “learn the landscape” (become familiar with authors, schools of thought, grasp the terminology, etc.), they later aspire to situate themselves (quoting one author rather than another, identification with a school of thought, adopting their own terminology). As such, while topics are useful for choosing a document or the parts of a thesis to be consulted, it is the meta-discursive elements that reveal the author’s stance which truly guide reading.

Our analysis of the corpus allowed us to identify the indicators of stance and interpret them. In doing so, the annotations added by readers and the oral comments associated with each passage of text allowed us to see how readers understood the documents they consulted. Such personal traces are a means for the reader to take possession of a document and interpret its content (Mille 2005). We analyzed 158 text fragments: of these, 129 had visual markers (underlining, highlighting, etc.); 47 contained annotations (notes, abbreviations, keywords, symbols); and 148 were commented on orally. The annotations and comments allowed us to identify two types of indicators in the fragments. The first occurred at the discourse level; the second at the textual level.

In the first case, the indicators collected were evaluative, axiological, and from epistemic and evidential categories. We, as such, found the linguistic markers mentioned in Grossman and Wirth (2010), Boch et al. (2007), and Rinck (2010), although there were fewer categories than in their research. In the second case, the indicators collected allowed us to localize statements according to their position in the document. We thus agree with Alain Berrendonner (1997) who has argued that “meta-discursive pointers” exist which are deictic (“here, see over”), text extracts (“in the first section”) or even imprecise locations (“in this passage”) and for whom a document is a “vectorized textual space.”⁷ To avoid all confusion between the

two sets of indicators, we prefer to talk about meta-discursive indicators when they help us find our way on the cognitive level and of meta-textual indicators when they help us find our way within the document.

4.0 Suggestions for indexing theses by stance

4.1 *Points of view, facets and terminological variations in stance?*

Unlike the notion of “point of view,” which finds resonance in information and documentation and amongst researchers in linguistics and computer science working on textual data (corpora, databases, the internet), the term stance is not commonly used in information science. In the context of information and documentation, indexing using Shiyali Ranganathan’s faceted classification system dates back to the 1950s. Facet analysis is not, strictly speaking, an enunciative approach that follows the author’s point of view, but rather it allows different points of view to be expressed about an object (Salvan 1962). Without reference to the famous classification system, Bachelin Ralalason (2010) has also employed the term facets when seeking to provide a multi-faceted representation of a document using several ontologies (ontology of topic, field, task, etc.). In this case, these representations involve the thematic content of a document, as well as its application context. Research conducted in the context of the RAP2 project has also underscored the interest of searching for information by point of view, thus allowing the user to focus on specific approaches to a concept. A whole collection of terms, called linguistic markers (Laublet et al. 2002), is associated with each point of view. To conclude this quick overview, let us mention research based on corpus linguistics which addresses scientific writing more specifically. The concept of point of view is central in pointing up an author’s scientific rhetoric (Teufel et al. 1999) and their enunciative position (Tutin et al. 2009) based on language. Such language markers are discontinuous, rooted in discourse or meta-discourse, and, as Ho-Dac and Péry-Woodley (2008, 3) have argued, they should not be confused with segmentation markers, but rather are indicators that “help nourish a relationship of continuity or discontinuity between two segments.”

4.2 *The triangular approach to stance*

Our previous research into the indicators of stance pointed up two important limitations: first, there is a

great diversity of markers that refer to numerous semantic categories which occasionally intersect and are difficult to grasp. Secondly, the dissemination of indicators throughout a text makes all attempts at indexing via this approach impossible. As such, we established that it is best to limit the notion of stance to three categories of markers that must simultaneously be found in a sentence or, at most, a paragraph.⁸ These categories set the triangular boundaries that delimit a stance's field of application: 1) Expressions that reveal a judgment or an author's subjective comments (agreement, mitigation, criticism, consensus, etc.); 2) expressions that name a topic (terms, concepts, propositional content, etc.); and 3) Expressions that mention the given environment (or give a reference mark)—this can be in discourse (dates, places, references to others, etc.) or in a document (chapter, section, etc.). Here are a few examples that contain indicators of stance. These extracts are part of a thesis read by one of the people interviewed for our research.

- E1. Il demeure cependant indéniable que l'hypertexte est un terme qui fait aujourd'hui partie de notre culture commune. (Ertzscheid, sujet 3)
- E2. Sans point commun apparent avec l'idée de Nelson, il est intéressant de remarquer comment, au point actuel de l'évolution technologique, les deux définitions entrent sans peine en résonance, laissant entrevoir un champ épistémologique à la fois ouvert et complexe dans lequel les associations de l'un font écho aux « dérives » de l'autre. (Ertzscheid, sujet 3)
- E3. Nous défendons dans ce travail la thèse selon laquelle l'hypertexte n'est pas un épiphénomène de nature informatique assimilable ou réductible à l'un des sphères de la réalité qui l'emploie. (Ertzscheid, sujet 3)

Each of these extracts contains the three categories of markers: 1) Expressions that identify a stance (assertion, statement, thesis): "il demeure indéniable que (E1), il est intéressant de remarquer que (E2), nous défendons la thèse selon laquelle" (E3); 2) Expressions that describe a topic: "hypertexte" (E1, E2, E3); and 3) Expressions that allow us to locate a point of view (time, place or angle of approach): "aujourd'hui" (E1), " sans point commun avec l'idée de Nelson" (E2) "En prenant l'angle critique qu'offre l'analyse des hypertextes littéraires" (E3) (document, chapter, imprecise location): "dans ce travail" (E3).

Although we do not yet have any precise data on the efficiency of this model, we have chosen (at first)

to recommend the most restrictive model since it requires that three levels of information be simultaneously present. Stance thus has composite status: it combines elements of language that are both indicial and relational and points to markers at different levels (lexical, syntactical). This proposal is based on the localist framework (see Clavier and Paganelli 2010).

4.3. Connection between indexing and practice

The systems that provide access to digitized theses offer various means to search for information: generally, access by structured field (author, title, etc.) and access by content (title, abstract or keyword).⁹ Occasionally, it is possible to search the entire text.¹⁰ In order to improve access to information in theses, we recommend including knowledge about authorial stance and connecting it to indexed topics. This representation would involve a twofold indexing process. After segmenting the text, each fragment from the cut-up would be described by both the topics it contains and a label indicating whether or not indicators of stance are present. Such dual indexing would exist on pre-identified and segmented units of information; we believe that paragraphs are the most appropriate basic units for the segmentation and indexing of large documents (Mounier and Paganelli 2003).

On the first level, topics would be indexed according to the structure of the document. This approach has notably been described by Abascal-Mena and Rumpler (2007) with regard to theses; an overview of existing methods for the thematic indexing of long documents like monographs has been done by Lyne Da Sylva (2004).

On the second level, units of information would be characterized according to whether or not indicators of stance are present. When indicators are present, the nature of the stance (critical, agreement, etc.) would be mentioned. The way indexes are structured offers for two possible solutions.

In the first case, indexes by topic and marker of stance would be dissociated; in the second case, one index would contain both sets of information: the topics and whether they do or do not contain stance markers. The first solution would be linguistically more coherent since there would be an index for each level of information. Conversely, the second solution would offer the advantage of listing topics that are or are not modalized. Both types of indexing would allow for research that combines searches by topic and stance; the indexes would need to be designed to be included in the primary document rather than be

separate; they would also need to be designed as reading tools that allow us to manipulate no longer entire documents but rather segments of text. In this respect, our proposals are similar to the recommendations made by Muriel Amar (2004) regarding the nature of indexes needed for new digital media.

5.0 Conclusion

We took as the basis for our research that different approaches to reading scientific documents could be interpreted through an analysis of the traces of informational activity. This methodology allowed us to empirically confirm the relevance of the notion of stance when consulting theses on the one hand and, on the other hand, the interest in associating an author's "global" point of view (criticism, agreement, consensus, etc.) towards the topics, notions, and concepts addressed in a document. We also suggested representing the infinite number of reading experiences in the form of stable knowledge likely to be represented in indexes. This research needs to be pursued with the systematic collection of markers in order to assess the degree of automation in indexing. This understanding of knowledge is related to indexing as an interpretive process that cannot be imposed by a controlled vocabulary or solely by the text but which is also mediated by the traces of an individual's use in the context of their work.

Notes

1. Cultural Identity in Academic Prose (2001-2005) directed by Kjersti Fløttum (University of Bergen)
2. Scientext: un corpus et des outils pour étudier le positionnement et le raisonnement de l'auteur dans les écrits scientifiques [a corpus and tools to study authorial stance and reasoning in scientific texts], directed by Francis Grossmann and Agnès Tutin, ANR 2007-2010 <http://scientext.msh-alpes.fr>
3. Scientext, *ibid.*, translated here.
4. In France, the information and communication sciences form a single discipline, which makes them somewhat of an exception.
5. From an internal document produced by the Grenoble sicd2: "the consultation figures for digital theses are impressive. For the 4000 theses available on the TEL/CCSD server, there are over 100 downloads per day, whereas a paper thesis is consulted on average once every ten years" (translated here).
6. Notably at the Lyon 2, Paris 12, Aix-Marseille 2 and Bordeaux 1 Universities and at Insa Lyon.
7. For Alain Berrendonner, a text is an "organized collection of successive utterances accomplished over the course of a discourse" and "vectorized space" is a "schematization of the text as space" [*ibid.* 221, translated here].
8. This last statement needs to be verified through more in-depth research.
9. TEL (tel.archives-ouvertes.fr/) or Fourier University's collection of online theses (tel-ujf.ujf-grenoble.fr/).
10. Lyon 2's "Cybertheses" (theses.univ-lyon2.fr) or theses from the Strasbourg universities (<http://scd-theses.u-strasbg.fr/>).

References

- Abascal-Mena, Rocio, and Rumpler, Béatrice. 2007. Accès au contenu des thèses numériques par leur structure sémantique *Document numérique 10*: 9-35.
- Amar, Muriel. 2004. L'indexation aujourd'hui. In *La fonction documentaire au coeur des TIC Les dossiers de l'ingénierie éducative 49*, décembre: 61-65.
- Bartlett, Joan C., and Toms, Elaine G. 2005. How is information used ? Applying task analysis to understanding information use. *Canadian Association for Information Science (CAIS) Proceedings*, The University of Western Ontario, 2-4 juin 2005, London, Retrieved January 27, 2012, from http://www.cais-acsi.ca/proceedings/2005/bartlett_2005.pdf
- Berrendonner, Alain. 1997. Schématisation et topographie imaginaire du discours, In Mieville, D. et Berrendonner, A. eds., *Logique, discours et pensée. Mélanges offerts à Jean-Blaise Grize*. Bern: Peter Lang, pp. 219-37.
- Boch, Françoise, Grossmann, Francis, and Rinck, Françoise. 2007. 'Conformément à nos attentes...', ou l'étude des marqueurs de convergence/divergence dans l'article de linguistique. *Revue Française de linguistique appliquée XII*: 109-22.
- Byström, Katriina. 2007. Approaches to task in contemporary information studies *Information research 12*(4), paper colis26, Retrieved January 27, 2012, from <http://informationr.net/ir/12-4/colis/colis26.html>
- Chaudiron, Stéphane, and Ihadjadene, Madjid. 2002. Quelle place pour l'utilisateur dans l'évaluation des SRI ? Recherches récentes en Sciences de l'Information, convergences et dynamiques. *Actes du colloque MICS-LERASS*. 21-22 mars 2002, Toulouse, Paris: ADBS Éditions, pp. 211-33.

- Cheuk, Wai-Yi B. 1999. The derivation of a situational information seeking and use process model in the workplace: employing sense-making. *International Communication Association annual meeting*, San Francisco, California, Retrieved January 27, 2012, from <http://communication.sbs.ohio-state.edu/sense-making/meet/1999/meet99cheuk.html>
- Clavier, Viviane, and Paganelli, Céline. 2010. Marqueurs de positionnement et parcours de lecture: un enjeu pour la consultation des thèses en ligne ? *Les enjeux de l'information et de la communication*, supplément 2010B, *Evolutions technologiques et information professionnelle: pratiques, acteurs et documents*, Retrieved January 27, 2012, from http://www.u-grenoble3.fr/les_enjeux
- Cotte, Dominique. 2004. Le concept de "document numérique", *Dossier Du "document numérique" au "textiel"* *Communication & langages* 140: 31-41.
- Couto, Javier and Minel, Jean-Luc. 2007. NaviTexte, a Text Navigation Tool. *AI*IA 2007, Artificial Intelligence and Human-Oriented Computing, Lecture Notes in Computer Science*, 4733/2007, pp.720-29.
- Da Sylva Lyne. 2004. Indexation automatique de documents par combinaison d'analyses statistiques et terminologiques structurées. *Actes du colloque RIAO'04*, Avignon, 26-28 avril 2004, pp.895-904.
- Fidel, Raya, and Pejtersen, Anneliese M. 2004. From information behaviour research to the design of information systems: the Cognitive Work Analysis framework. *Information research* 10(1), paper 210, Retrieved January 27, 2012, from <http://InformationR.net/ir/10-1/paper210.html>
- Flon, Émilie, Davallon, Jean, Tardy, Cécile, and Jeanret, Yves. 2009. Traces d'écriture, traces de pratiques, traces d'identité, *Actes du colloque H2PTM'09*, 3 septembre, 1 et 2 octobre 2009, *Rétrospective et prospective 1989-2009*, Paris: Hermès - Lavoisier, pp. 181-92.
- Fondin, Hubert. 2001. La science de l'information: posture épistémologique et spécificité disciplinaire. *Documentaliste sciences de l'information* 38: 112-22.
- Fondin, Hubert. 2005. La science de l'information ou le poids de l'histoire. *Les Enjeux de l'information et de la communication*, 2005. Retrieved January 27, 2012, from http://w3.u-grenoble3.fr/les_enjeux/2005/Fondin/index.php
- Guyot, Brigitte. 2002. Mettre en ordre les activités d'information, nouvelle forme de rationalisation organisationnelle. *Les enjeux de l'information et de la communication*, 2002. Retrieved January 27, 2012, from http://www.u-grenoble3.fr/les_enjeux
- Hochon, Jean-Claude, and Jacoboni, Eric. 1994. Lecture active et annotation dynamique assistées par ordinateur. In Tankoano J. Ed., *Actes du colloque africain sur la recherche en informatique*. Colloques et Séminaires. Paris: ORSTOM, pp. 509-26.
- Ho-Dac, Lydia-Mai, Péry-Woodley, Marie-Paule. 2008. Méthodologie exploratoire outillée pour l'étude de l'organisation du discours, in Durand J., Habert, B. & Laks B. Eds., *Congrès Mondial de Linguistique Française (CMLF-08)*, Paris, 9-12 juillet 2008.
- Huvila, Isto. 2008. Work and work roles: a context of tasks *Journal of documentation* 6: 797-815.
- Järvelin Kalervo, Ingwersen Peter. 2004. Information Seeking Research Needs Extension toward Tasks and Technology *Information Research*, 10(1), paper 212, Retrieved January 27, 2012, from <http://informationr.net/ir/10-1/paper212.html>
- Kuhlthau Carol. C. 2004. *Seeking meaning: a process approach to library and information services*, 2nd edition. Westport: Libraries Unlimited.
- Lainé-Cruzet, Sylvie. 1999. PROFILDOC: filtrer une information exploitable. *Bulletin des bibliothèques de France* 44/5: 60-64.
- Laublet, Philippe, Nait-Baha, Laublet, Jackiewicz, Agata, and Djioua, Brahim. 2002. Collecte d'informations textuelles sur le Web selon différents points de vue. In Paganelli, Céline, ed., *Interaction homme-machine et recherche d'information*. Paris: Hermès.
- Le Coadic. Yves-François. 1994. *La science de l'information*. Que sais-je ?, 2873. Paris: PUF.
- Li, Yuelin, and Belkin, Nicholas J. 2008. A faceted approach to conceptualizing tasks in information seeking. *Information processing & management* 44:1822-37.
- Mille, Dominique. 2005. *Modèles et outils logiciels pour l'annotation sémantique de documents pédagogiques*, thèse de doctorat en Informatique, Université Grenoble 1, Retrieved January 27, 2012, from <http://tel.archives-ouvertes.fr/tel-00011468>
- Miranda, Sylvania V. and Tarapanoff, Kira M.A. 2007. Information needs and information competencies: a case study of the off-site supervision of financial institutions in Brazil. *Information research* 13(2), paper 344, Retrieved January 27, 2012, from <http://informationr.net/ir/13-2/paper344.html>
- Mounier, Evelyne, and Paganelli, Céline. 2003. Segmentation d'un document en paragraphes: une application à la recherche d'information dans les documents techniques très volumineux. *Modèles linguistiques* XXIV, fasc. 2: 85-98.

- Paganelli, Céline and Mounier, Evelyne. 2002. Vers un système de consultation des documents techniques volumineux par des utilisateurs experts: le système Sysrit. In Paganelli Céline, ed., *Interaction homme-machine et recherche d'information*. Paris: Hermès, pp.195-228.
- Paganelli, Céline and Mounier, Evelyne. 2009. Stratégies informationnelles en milieu professionnel: du réseau personnel à la Toile, In *congrès international ACSI (association canadienne des sciences de l'information). Tracer l'horizon informationnel du XXIe siècle: Frontières, passerelles et carrefours*, 28-30 mai 2009, Ottawa, 18p. http://www.cais-acsi.ca/proceedings/2009/Paganelli_Mounier_2009.pdf
- Péry-Woodley, Marie-Paule, and Scott, Donia. 2006. Computational Approaches to Discourse and Document Processing *T.A.L.* 47(2): 7-19.
- Polity, Yolla. 2000. L'évolution des paradigmes dans le domaine de la recherche d'information. *Communication au groupe de travail Théories et Pratiques Scientifiques (TPS) de la SFSIC*, 3 mars 2000. Retrieved January 27, 2012, from http://www.iut2.upmf-grenoble.fr/RI3/TPS_paradigmes.htm
- Poudat, Céline, Cleuziou, Guillaume, and Clavier, Viviane. 2006. Catégorisation de textes en domaines et genres: complémentarité des indexations lexicales et morphosyntaxiques *Document numérique* 9/1: 61-76.
- Ralalason, Bachelin. 2010. Représentation multifacette des documents pour leur accès sémantique. Thèse de doctorat en informatique, Université de Toulouse 3, Retrieved January 27, 2012, from http://thesesups.ups-tlse.fr/1050/1/Ralalason_Bachelin.pdf
- Rinck, Fanny. 2010. L'analyse linguistique des enjeux de connaissance dans le discours scientifique. Un état des lieux. *Revue d'anthropologie des connaissances* 4/3: 427-50.
- Salvan, Paule. 1962. Les progrès de la classification à facettes. *Bulletin des bibliothèques de France* 7/1: 5-17.
- Staii, Adrian, Balicco, Laurence, Bertier, Marc, Clavier, Viviane, Mounier, Evelyne, and Paganelli, Céline. 2006. Les pratiques informationnelles des médecins dans les centres hospitaliers universitaires: au croisement de la logique scientifique et de la culture professionnelle. *Revue canadienne des sciences de l'information et de bibliothéconomie* 30: 69-90.
- Teufel, Simone, Carletta, Jean, and Moens, Marc. 1999. An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of EACL Utrecht, Netherlands*, pp. 110-17.
- Tutin, Agnès, Grossmann, Francis, Falaise, Achille, and Kraif, Olivier. 2009. Autour du projet Scientext: étude des marques linguistiques du positionnement de l'auteur dans les écrits scientifiques. *Linguistique de corpus*, 2009, Lorient, Retrieved January 27, 2012, from http://w3.u-grenoble3.fr/lidilem/labo/file/Lorient_vfinale.pdf
- Yahiaoui, Leila, Prié, Yannick, Champin, Pierre-Antoine, and Boufaïda, Zizette. 2011. Redocumenting computer mediated activity from its traces: a model-based approach for narrative construction. *Journal of digital information* 12, retrieved January 27, 2012, from <http://journals.tdl.org/jodi/article/view/2088/1756>