Manuel Zacklad
Conservatoire National des Arts et Métiers, Laboratoire DICEN EA 4420, France[1]

Les écritures de l'information dans les TIC et le SI

Résumé : Nous présentons un cadre d'analyse du système d'information comme relevant essentiellement de procédés d'écriture variés empruntant à quatre catégories principales : le format de l'écriture (libre ou codifié), le régime de description (primaire ou secondaire), la signification (référentielle ou interprétative), le statut dans le document (principal ou auxiliaire).

Abstract: This paper presents an information system analysis framework as being essentially dependent on diverse writing techniques borrowing from four main categories: the writing format (open-ended or coded), the level of description (primary or secondary), the meaning (referential or interpretative), and its status within the document (main or associated).

1. Introduction

Dans cette présentation nous exposerons un cadre d'analyse visant à rendre compte du fonctionnement du système d'information comme relevant essentiellement de procédés d'écriture variés visant à médiatiser des transactions coopératives entre des personnes (cf. le cadre d'analyse de la sémiotique des transactions coopératives (Zacklad, 2008). Une grande partie de ces activités d'écriture relèvent d'une forme particulière, l'écriture codifiée, qui permet une automatisation partielle à la fois de l'écriture elle-même et de l'exploitation des écrits automatisés à l'aide de divers algorithmes de « requête » ou de « recherche ». Ce cadre vise à permettre de rendre compte de manière unifiée de la diversité croissante des technologies actuellement utilisées pour la description et la recherche des informations consignées (liées à des supports documentaires). Si les technologies sont souvent perçues à travers le prisme de la numérisation des supports, elles impliquent simultanément des techniques d'écriture.

Parmi celles-ci, les écritures codifiées qui représentent à l'aide d'expressions figées plus ou moins arbitraires le contenu d'un document, connaissent une importance croissante : langages documentaires (thésaurus, langages de classification…), langages de représentation des connaissances (réseaux sémantiques, cartes de thèmes, ontologies formelles et sémiotiques, etc.), index des moteurs de recherche créés par des algorithmes informatiques.

[1] Nous remercions pour son soutien le projet ANR Miipa-Doc (Méthodes et Services Intégrés Institutionnels et Participatifs pour la Classification à Facettes des Contenus Documentaires Complexes n°2008 CORD 014 03).

2. Format : écriture libre versus écriture codifiée (manuelle et automatique)
La première distinction que nous introduirons sépare l'écriture libre de l'écriture codifiée. La première et la plus commune dans la communication écrite est l'écriture « libre ». Le second format est la *codification.* Au niveau le plus simple, elle consiste à utiliser une expression figée qui devient un idiomatisme (au sens de Rastier, 1997) pour décrire une situation d'activité ou un contenu. La codification peut être manuelle ou automatique. La codification automatique s'appuie sur des procédés de *mesure automatique* dans le cas ou l'objet décrit est une situation non préalablement documentée (cf. infra) et sur des algorithmes dits de « fouille » dans le cas de l'analyse de contenu. Initialement limités aux textes, ceux-ci s'étendent aujourd'hui aux images et aux sons.

L'écriture codificatrice peut puiser dans un langage de codification préexistant sans rajouter de nouveau termes ou de nouveaux codes (codification fermée) ou créer de nouvelles propositions de termes ou de codes au fur et à mesure que l'analyse de la situation ou du contenu progresse (codification ouverte). Dans ce second cas, quand la codification est manuelle, se pose la question de la portée et des critères d'évaluation de la pertinence de la codification proposée.

3. Description primaire ou description secondaire
La seconde distinction que nous introduirons sépare les écrits selon qu'ils rendent compte d'une situation externe ou du contenu d'un document préexistant (que l'écriture soit libre ou codifiée). Le premier type d'écrit (que nous caractérisions comme produisant une *modélisation* dans Zacklad et al., 2008 et Zacklad, 2010) vise à représenter de manière schématique les composantes d'une situation transactionnelle qui n'a pas été préalablement documentée, c'est-à-dire qui n'a pas donnée lieu à la réalisation préalable d'une production sémiotique consignée via une transcription ou un enregistrement (texte, image, source sonore). Par contraste, quand la référence de l'écriture est un contenu déjà déposé dans un document, l'écrit correspond à une « analyse de contenu ».

4. Signification : relation référentielle ou interprétative de l'écrit à son l'objet
Nous distinguerons ensuite deux principaux types de relation de l'écriture à son objet ou deux types de signification : référentielle et rhétorico-herméneutique (terme emprunté à Rastier, 2003) ou interprétative[i] dont nous avons explicité les spécificités dans des publications relatives au web socio-sémantique et à la méta-sémiotique Hypertopic (Zacklad 2005, Zacklad et al. 2007, 2008). Dans (Zacklad, 2005), pour préciser le sens de ce terme, nous nous appuyons sur les travaux de (Mounin, 1968), qui propose une subdivision du signifié selon trois tendances relevant d'autant de théories : *logiques, contextuelles, situationnelles* (ou pragmatiques). Si l'on restreint la signification du terme de *logique* au domaine de la logique mathématique, on peut alors considérer que la première approche du signifié correspond à la signification référentielle. Celle-ci aborde la signification sur la base d'une correspondance entre l'expression et des propriétés physiques de la situation transactionnelle qu'elle décrit. De ce fait, la signification est indépendante à la fois du contexte de production de l'expression, les autres productions sémiotiques qui l'environnent ou les autres éléments de contenu du document (p.e. le texte dont l'expression fait partie) que de la situation de production elle-même, artefacts médiateurs, personnes, environnement.

Contexte et situation correspondent aux deux facettes herméneutiques et rhétoriques de la sémantique rhétorico-herméneutique ou interprétative. Le contexte correspond au texte ou aux éléments de contenu qui environnent l'expression. Ces éléments contextuels, à

commencer par le genre du document, connotent le sens de l'expression qui ne peut donc pas être appréhendé indépendamment de celui-ci. Cela a pour conséquence que la même expression, utilisée dans deux contextes différents, a également une signification en partie différente, mais cela signifie également que le contexte est une ressource qui permet d'expliciter, de désambigüiser, la signification de l'expression. Le contexte peut même être élargi au support (support papier ou électronique dans le cas d'un texte) ce qui implique qu'en toute rigueur, il est impossible de postuler une réelle indépendance du contenu et du support, malgré les affirmations dans ce sens qui font florès depuis la généralisation du numérique.

5. Statut par rapport au document : contenu principal versus d'un contenu auxiliaire (annotation et documentarisation)

Enfin, une quatrième distinction sépare l'écriture des éléments de contenu principaux d'un document de l'écriture des annotations qui contribuent à en faciliter la lecture et la gestion. Selon nous (Zacklad, 2007a), la retranscription ou l'enregistrement d'un contenu sur un support pérenne ne suffit pas à constituer celui-ci en document. Il faut pour cela « documentariser » le support, c'est-à-dire le doter d'annotations qui faciliteront son exploitation et sa ré-exploitation : mise en forme typographique, titrage et sous-titrage, indexation, ajout de métadonnées, etc.

6. Quatre dimensions pour définir la sémantique de l'information dans les TIC et le SI

Ces dimensions nous permettent ainsi de désambigüiser la sémantique des écrits et plus particulièrement des écrits numériques produits pour alimenter le système d'information et le Web. Que les écrits soient manuels ou automatiques nous proposons de les qualifier en répondant aux quatre questions suivantes. Les écrits étant de l'information consignée et chaque réponse étant binaire, notre cadre d'analyse détermine seize catégories sémantiques pour la caractérisation de l'information.

- Le format de l'information est-il libre ou codifiée (format) ?
- L'information correspond elle à une description primaire (modéliser d'une situation) ou secondaire (régime de description) ?
- La signification de l'information est-elle référentielle ou interprétative (signification) ?
- L'information constitue-t-elle le contenu principal du document ou a-t-elle un statut auxiliaire (statut) ?

Nous présenterons dans un tableau un exemple d'écrits pour chacune de ces seize catégories, dont voici quelques exemples :

Dimension **Type d'écrit**	Format : libre ou codifiée	Régime de description : primaire vs secondaire	Signification : référentielle ou interprétative	Statut dans le document : contenu principal ou auxiliaire
Article de journal décrivant un fait divers	Format : libre	Description : primaire	Signification : référentielle	Statut : principal
Critique de film	Format :	Description :	Signification :	Statut : principal

dans un journal	libre	secondaire	interprétative	
Notice d'un livre en bibliothéconomie	Format : libre	Description : secondaire	Signification : interprétative	Statut : auxiliaire
Métadonnée décrivant l'éditeur du document	Format : codifié	Description : primaire	Signification : référentielle	Statut : auxiliaire
Terme de thésaurus décrivant un document	Format : codifié	Description : secondaire	Signification : interprétative[ii]	Statut : auxiliaire
Terme d'un index d'un moteur de recherche automatique	Format : codifié	Description : secondaire	Signification : référentielle	Statut : auxiliaire
Concept d'une ontologie formelle	Format : codifié	Description : primaire	Signification : référentielle	Statut : auxiliaire ou principal selon l'usage
Description d'un article dans une base de données	Format : codifié	Description : primaire	Signification : référentielle	Statut : principal

Fig.1. Exemple d'écritures de l'information relevant des différentes catégories

7. Bibliographie

Amar, Muriel. 2000. *Les fondements de l'indexation, une approche linguistique*, ADBS Edition, Paris.

Marleau, Yves, Mas Sabine, Zacklad Manuel. 2008. Exploitation des facettes et des ontologies sémiotiques pour la gestion documentaire. In *Traitements et pratiques documentaires : vers un changement de paradigme ?* eds. Broudoux, Evelyne & Chartron, Ghislaine. Paris : ADBS Editions, p. 91-110.

Mounin, Georges.1968. *Clefs pour la linguistique*, Seghers, Paris.

Pédauque, R. T. 2006. *Le document à la lumière du numérique*, Caen : C&F Editions.

Rastier, François. 1996. La sémantique des textes : concepts et applications, *Hermès*, N°16, pp. 15-37.

Rastier, François.1997. Défigements sémantiques en contexte. In *La locution, entre langues et usages*, ed. Martins-Baltar, M., coll. Signes, ENS Editions Fontenay / Saint Cloud, diff. Ophrys, Paris, 1997, pp. 305-329. En ligne [http://www.revue-texto.net/Inedits/Rastier/Rastier_Defigements.html]

Rastier, François. 2003. De la signification au sens. Pour une sémiotique sans ontologie. *Texto!* juin-sept. 2003 en ligne [http://www.revue-texto.net/Inedits/Rastier/Rastier_Semiotique-ontologie.html]

Zacklad, Manuel. 2005. Introduction aux ontologies sémiotiques dans le Web Socio Sémantique. In Ed. Jaulent, M. *IC. 16èmes journées francophones d'Ingénierie des Connaissances*, 30-03 Avril 2005, Nice. Grenoble: PUG, 12 p.

Zacklad, Manuel. 2007(a). Processus de documentarisation dans les Documents pour l'Action (DopA). In Ed. Babel. *Le numérique : impact sur le cycle de vie du document*. ENSSIB - février 2007 [en ligne] http://babel.enssib.fr/document.php?id=727

Zacklad, Manuel. 2007(b). Annotation : attention, association, contribution. In Salembier P. et Zacklad M. eds. *Annotations dans les Documents pour l'Action*. Lavoisier, Paris : 29-46.

Zacklad, Manuel. 2007(d), Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information, in *Actes de la Conférence CAIS/ACSI* 2007, Montréal, http://www.cais-acsi.ca/search_fr.asp?year=2007

Zacklad, Manuel. 2008. Ordre du discours : cadrage et recadrage des transactions communicationnelles dans les processus de changement et d'innovation au sein des organisations, In *Actes du colloque « Qu'est-ce qu'une organisation ? Matérialité, action et discours »*, 21 et 22 mai, Montréal.

Zacklad, Manuel., Cahier, Jean-Pierre., Bénel, Aurélien., Zaher, Hédi., Lejeune, Christophe., Zhou, Chao. 2007. Hypertopic : une métasémiotique et un protocole pour le Web socio-sémantique, in *Actes des 18èmes journées francophones d'Ingénierie des Connaissances*, 2-6 juillet 2007, Grenoble, en ligne [http://archivesic.ccsd.cnrs.fr/sic_00180793/en/]

Zacklad, Manuel., Cahier, Jean-Pierre., Bénel, Aurélien., Zaher, Hédi., Lejeune, Christophe., Zhou, Chao. 2008. Intelligence qualitative dans la gestion de risque : rôle de la documentarisation et des Systèmes d'Organisation des Connaissances, in Prié, Y. Ed. *IC 2008: Ingénierie des connaissances 2008*, INPL, en ligne [http://orbi.ulg.ac.be/handle/2268/19612]

---

[i] Pour simplifier la formulation nous parlons ici de signification interprétative pour décrire ce que nous nommons sémantique rhétorico-herméneutique dans nos autres textes. La sémantique interprétative est la théorie générale de la signification proposée par F. Rastier qui s'inscrit dans un paradigme différentialiste du sens (opposé à des paradigmes référentiels et inférenciels, cf. Duteil-Mougel 2004)

[ii] Le statut interprétatif ou non de l'indexation doit faire l'objet d'une problématisation (cf. Amar 2000).