



HAL
open science

Processus de documentarisation dans les Documents pour l'Action (DopA) : statut des annotations et technologies de la coopération associées

Manuel Zacklad

► To cite this version:

Manuel Zacklad. Processus de documentarisation dans les Documents pour l'Action (DopA) : statut des annotations et technologies de la coopération associées. Oct 2004. sic_00001072v1

HAL Id: sic_00001072

https://archivesic.ccsd.cnrs.fr/sic_00001072v1

Submitted on 23 Sep 2004 (v1), last revised 15 May 2005 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Processus de documentarisation dans les Documents pour l'Action (DopA) : statut des annotations et technologies de la coopération associées

Manuel Zacklad

Manuel.Zacklad@utt.fr
Université de Technologie de Troyes
Equipe Tech-CICO / ISTIT FRE CNRS 2732

Résumé :

Dans cette communication, nous nous intéresserons particulièrement aux situations dans lesquelles les documents sont des supports à la coordination d'un collectif distribuée engagé dans une activité commune finalisée. Après avoir défini la notion de *production sémiotique* résultant d'une *transaction communicationnelle* nous introduirons les stratégies permettant une *distribution spatio-socio-temporelle* de ces transactions. Parmi ces stratégies, nous étudierons plus en détail la stratégie de *documentarisation* qui consiste à pérenniser le support matériel de la transaction et à le doter d'attributs permettant sa ré-exploitation. L'étude des processus de documentarisation nous conduira à introduire plusieurs notions : la distinction *transcription vs enregistrement*, la *double articulation documentaire* externe vs interne du document, les *liaisons mécanique et organique* entre les fragments du document. Ces notions nous permettront de définir précisément les *Document pour l'Action* (DopA) et leurs caractéristiques. L'activité d'annotation sera définie comme une activité permettant d'articuler des fragments documentaires sur des DopA pour soutenir efficacement les activités coopératives. Cette vision nous permettra d'analyser un certain nombre de technologies de la coopération et en particulier les forums de discussion comme relevant de technologies annotatives. Nous concluons en soulignant l'importance de la codification dans les processus d'investissement documentaire et les conditions permettant d'y recourir.

Mots-clefs : Document, Document pour l'Action (DopA), Documentarisation, Annotation, Rédaction Coopérative, Forum de discussion

1 Evolution du document numérique : un point de vue économique-gestionnaire, CSCW et ingénierie des connaissances

L'essor du Web dans un contexte internet ou intranet a entraîné une multiplication des pratiques collectives centrées sur le document que celles-ci soient de nature strictement professionnelles (conception médiatisée par des plans et des contrats, prise en charge médicale ou médico-sociale par un réseau de santé, par exemple), ou qu'elles soient moins structurées (activité d'associations militantes ou de communauté d'intérêt partageant une passion commune). Toute une série de nouvelles pratiques centrées sur le document sont ainsi entrain d'émerger, allant de la mise à disposition de documents dans des petits systèmes de gestion documentaire amenés à constituer des « micro bibliothèques numériques » partagées par une équipe à l'utilisation des annotations accompagnant la circulation ou la rédaction collective du document en passant par l'usage des Wiki et Blog.

Pour rendre compte de la multiplicité de ces pratiques et du caractère à la fois collectif et évolutif des documents ou collections en cours de rédaction, nous avons proposé de définir la notion de Document pour l'Action (DopA, Zacklad 2003c). Ceux-ci possèdent un certain nombre de caractéristiques (inachèvement prolongé, pérennité, fragmentation, rapport non trivial aux auteurs, rapport non trivial aux éléments du signifié...) qui questionnent à plusieurs titres les nouvelles « théories du document » et notamment la problématique de la gestion de son cycle de vie compte tenu du fait que son existence publique peut commencer avant qu'il soit achevé et qu'il est en mesure de rassembler de nouvelles contributions sous la forme d'annotations après avoir été stabilisé dans une première version.

Dans cette communication, nous nous intéresserons particulièrement aux situations dans lesquelles les documents sont des supports à la coordination d'un collectif distribuée engagé dans une activité commune finalisée. Dans ces contextes, le document apparaît comme un ensemble de fragments portés par des auteurs divers dont le contenu final reste largement indéterminé alors même que sa circulation rapide lui fait déjà jouer un rôle majeur d'information, d'aide à la décision et de preuve. Ce statut particulier des DopA interpelle au premier chef la problématique de l'indexation : le caractère évolutif et fragmentaire des DopA et de leurs annotations justifie la mise en place de nouveaux dispositifs de codification que nous évoquerons notamment en conclusion.

Nous suivrons le plan suivant. Après avoir défini la notion de production sémiotique résultant d'une transaction communicationnelle (Zacklad 2004) entre un ou plusieurs réalisateurs et un ou plusieurs bénéficiaires nous introduirons la problématique de la distribution spatio-socio-temporelle de ces transactions qui implique la mise en place de stratégies permettant leur préservation. Parmi ces stratégies, au nombre de quatre, nous étudierons plus en détail la stratégie de documentarisation qui consiste à pérenniser le support matériel de la transaction et à le doter d'attributs permettant sa ré-exploitation. L'étude des processus de documentarisation nous conduira à introduire plusieurs notions : la distinction *transcription vs enregistrement*, la *double articulation documentaire* externe vs interne du document, les *liaisons mécanique et organique* entre les fragments du document.

Ces notions nous permettront de définir précisément les DopA et leurs caractéristiques. Nous verrons en particulier qu'ils constituent un dispositif pour la mise en place de *transactions communicationnelles émergentes distribuées*. L'activité d'annotation sera définie comme une activité permettant d'articuler des fragments documentaires sur des DopA et donc comme leur permettant de soutenir efficacement les activités coopératives. Cette vision nous permettra d'analyser un certain nombre de technologies de la coopération et en particulier les forums de discussion comme relevant de technologies annotatives. Nous concluons en soulignant l'importance de la codification dans les processus d'investissement documentaire, particulièrement pour les documents numériques, et les conditions permettant d'y recourir.

2 Une définition communicationnelle et pragmatique des concepts de document et de production sémiotique

2.1 Hyperédaction, hyperfragmentation

Le point de vue spécifique qui est le notre élargit la perspective documentaire bien que celle-ci se soit déjà, à travers bon nombre de ses auteurs (cf. par exemple, Brown & Duguid 1995, Briet 1951, Buckland 1997, Pédaque 2003), émancipée d'un rapport trop exclusif aux textes

« classiques ». Elle le fait en s'ancrant résolument dans une perspective communicationnelle et pragmatique, elle-même appuyée sur le cadre théorique des transactions communicationnelles symboliques (Zacklad 2004), qui vise à rendre compte des activités coopératives, notamment médiatisées par divers artefacts symboliques, dans une perspective à la fois communicationnelle, cognitive et socio-économique.

Si nos réflexions visaient clairement à l'origine des documents d'abord produits et diffusés au sein d'organisations professionnelles dont l'activité est assez précisément finalisée et ce dans le contexte spécifique de la numérisation généralisée qui permet des relations fluides entre intranet et internet, nous pensons qu'elles peuvent également s'appliquer à l'ensemble des formes de documents jusqu'alors traités par les sciences de l'information, dont les finalités sont le plus souvent « culturelles » au sens large.

Pour spécifier les systèmes d'information de nature documentaire qui véhiculent les documents en permettant leur conception (p.e. rédaction) et leur exploitation (p.e. lecture) et qui correspondent aux usages très divers d'internet au début du troisième millénaire (circulation accélérée des fichiers via la messagerie, gestionnaire de contenu pour les sites web, forum de discussion, annotation en ligne, weblogs...) il nous a semblé nécessaire de contribuer à une redéfinition du *concept de document* et de manière corollaire des concepts qui appréhendent *les contenus* de ces documents qui relèvent de moins en moins de la catégorie du texte classique pour inclure des images et des sons accessibles à partir de systèmes de navigation de nature de plus en plus variée.

Mais surtout, avec la progressive maturation des pratiques d'internet, on assiste à une multiplication des usages collectifs des documents numériques, notamment lors de leur phase de production. Les conséquences de ces processus de conception collective et en particulier d'une conception prolongée sur la durée, pour l'appréhension du concept de document, sont sans doute aussi importantes qu'a pu l'être en son temps la prise en compte d'objets tangibles dans le champ des réflexions bibliographiques (Otlet 1934, Briet 1951) puis celle, plus récente, induite par la prise en compte de la numérisation généralisée des supports et les modalités de navigation hypertextuelle associées (Buckland 1998, Brown & Duguid 1995, Pédaque 2003). En risquant un nouveau néologisme, nous considérerons qu'après la phase d'hyper-lecture rendue possible par les hypertextes, succède une phase d'hyper-rédaction dans laquelle les documents pour l'action numérisés apparaissent eux-mêmes comme constitués d'une multiplicité d'*hyperfragments documentaires mouvants* dont l'unification dépend des activités collectives qu'ils visent à soutenir.

2.2 Production sémiotique associée à une transaction communicationnelle

Toute analyse de la notion de document nous semble devoir inscrire ceux-ci dans un processus de communication pour partie différé, au sens des processus asynchrones décrits dans le champ du CSCW¹, entre des producteurs et des récepteurs liés par des intérêts communs. Pour l'exprimer dans les termes de la théorie des transactions communicationnelles symboliques, le document est analysé comme étant l'objet d'une transaction entre des acteurs impliqués dans un processus d'échange visant à la fois des engagements mettant en jeu leur « self » et des connaissances liées à la production d'une « œuvre » au moins pour partie commune (Zacklad 2004).

¹ Computer Supported Cooperative Work ou TCAO, Travail Coopératif Assisté par Ordinateur.

Comme nous le décrivions dans le contexte plus restreint de la « théorie des transactions intellectuelles » (Zacklad 2000), les transactions médiatisées par des objets documentaires peuvent apparaître comme des *pseudo-transactions artéfactualisées* au sein desquelles le support de la transaction communicationnelle est de nature pérenne. La transaction peut concerner des acteurs singuliers ou, dans certains contextes, viser une audience plus large dont les membres ne sont pas dénommés individuellement, la transaction prenant alors un caractère universalisant. Les documents associés, qui étaient alors exclusivement analysés comme concernant des écrits, pouvaient être qualifiés d'écrits chauds, tièdes ou froids, selon que les engagements associés à la transaction conservaient ou non un caractère actif².

Mais la forme documentaire n'est qu'une des matérialisations possibles de l'objet de la transaction. Pour analyser plus précisément les spécificités du document parmi l'ensemble des objets sémiotiques qui participent des procès de communication, il est nécessaire de se replacer dans le contexte plus large de l'analyse du fonctionnement des transactions communicationnelles dont les supports peuvent être aussi variés que la posture, le geste ou la voix pour ne citer que les plus fréquents. Dans ce contexte élargi, nous considérerons que l'objet de la transaction est une *production sémiotique* élaborée par un producteur pour un bénéficiaire dans une *situation transactionnelle* que nous décrivons ici spécifiquement sous un angle communicationnel³.

Les composantes de la *situation transactionnelle* (qui peut elle-même se décomposer dans certains cas particuliers en situation de production sémiotique et situation de réception) comprennent :

- un ou plusieurs *réalisateurs* ;
- un ou plusieurs *bénéficiaires*⁴ ;
- des *paramètres de la situation transactionnelle* au sens où ce terme a pu être utilisé dans le domaine de l'analyse pragmatique de l'énonciation (Ducrot 1984), qui relèvent des domaines suivants :
 - un *projet commun* sur la base d'intérêts ou d'objectifs pour partie partagés qui justifient la participation à cette situation collective ;
 - des *relations sociales* entre le réalisateur et le bénéficiaire dont découlent pour partie les intérêts ou objectifs communs poursuivis ;
 - un *cadre spatio-temporel* et des *conditions environnementales* de nature non strictement sociales susceptibles d'influencer tant l'atteinte des objectifs que le processus de production sémiotique ;
 - les contraintes pesant sur le choix du *média* qui se décompose lui-même dans le choix d'une *modalité d'expression* (langagière orale, scripturale, gestuelle, filmique, etc...) et sur le choix d'un *support matériel* associé à ce média quand plusieurs options sont possibles (modalités de vocalisation, nature du support pour un média scriptural, etc...) ;

² On notera que le caractère plus ou moins actif de la transaction dont témoigne le document n'est pas liée à son ancienneté comme le montre l'extrême vivacité du rapport aux textes dit « sacrés » pour certaines communautés.

³ Le fait de se focaliser ici sur la dimension sémiotique des objets produits lors des transactions ne doit pas faire oublier le fait qu'il ont également d'autres dimensions, économique ou physique, par exemple.

⁴ Il ne nous est pas possible de développer ici les raisons pour lesquels nous utilisons les concepts de « producteur » et de « bénéficiaire » au lieu, par exemple, des concepts « d'émetteur » et de « récepteur » utilisés dans le modèle shanonien classique. Signalons simplement que la vision mathématique de l'information à laquelle se réfère ce modèle est très éloigné de la vision anthropologique des objets sémiotiques sur laquelle nous nous basons. Quand le terme de « réception » est utilisé pour décrire les situations où les bénéficiaires prennent connaissance du média et de son contenu il se rapproche d'avantage du concept de « consommation » que de celui « d'enregistrement » au sens informatique du terme.

- un *terrain représentationnel commun* correspondant à l'ensemble des représentations communes évoquées issues des connaissances communes associées à la situation transactionnelle et à l'histoire de la transaction, terrain commun indispensable à l'intercompréhension (voir la note 4 sur la notion de représentations communes par rapport aux notions de croyances communes, connaissances communes, etc.) ;
- les *compétences* des réalisateur(s) et bénéficiaire(s) liées à la situation au sens large et à l'activité communicationnelle leur permettant pour les uns, de réaliser une production sémiotique sur un média donné et pour les autres, de recevoir ce média et cette production ;
- une *production sémiotique* véhiculant un *contenu sémiotique* via un *média* élaborée par les réalisateurs à l'intention des bénéficiaires.

Dans l'analyse de la production sémiotique, on pourra établir une distinction entre d'une part la forme et le support de cette production, ou *média de la production sémiotique* et d'autre part le *contenu sémiotique* « véhiculé » par ce média. On verra dans la suite de l'analyse que nous considérons les documents comme des productions sémiotiques médiatisées par des supports pérennes dotés de propriétés spécifiques.

La production sémiotique se compose donc :

- du *média* qui peut lui-même être analysé, comme on l'a vu plus haut, selon deux paramètres :
 - le choix d'une *modalité d'expression* (langagière orale, scripturale, gestuelle, filmique, etc...), au sein de laquelle est conventionnellement déterminé un système de signes permettant des agencements syntagmatiques et paradigmatiques ;
 - le choix d'un *support matériel* associé à cette modalité d'expression (type de vocalisation, amplitude des gestes, nature du support pour un média scriptural, etc...) adapté aux caractéristiques de la modalité d'expression ;
- du *contenu sémiotique* (ou signification) qui peut lui-même être appréhendé de deux manières en fonction de :
 - son *pouvoir d'évocation* qui est sa capacité à évoquer des *représentations communes*⁵ en fonction (i) de l'agencement des signes conforme aux possibilités offertes par le média et (ii) des paramètres de la situation transactionnelle (le *pouvoir d'évocation* correspond à la facette traditionnellement étudiée par la sémantique). Dans certaines situations extrêmement standardisées les signes n'évoquent pas de représentation mais déclenchent directement des « automatismes » et l'on considérera que les

⁵ La notion de représentation commune correspond pour nous à la notion de « terrain commun » au sens de Clark 1996. Selon Clark (cf. p 94), qui s'appuie sur Lewis 1969, la notion de terrain commun correspond à « l'information possédée » (has information) par les participants, expression qui subsume les catégories de croire, connaître (know), assumer, ou être conscient, qui elles-mêmes renvoient aux notions techniques de croyance mutuelle, connaissance mutuelle, assomption mutuelle, et conscience mutuelle. La notion de « représentation commune », plus proche de la psychologie cognitive, à pour nous l'avantage de faire référence aux processus cognitifs d'activation reconstructive de l'information dans le cadre de transactions communicationnelle. La notion de « connaissance commune » renverrait d'avantage pour nous à la notion de « communal common ground », d'information partagée par des communautés élargies garantes de leur pérennité notamment par le biais de différents artefacts (le terme « d'information » est neutre pour nous et renvoie à la « sémantique » d'une production sémiotique -ie les représentations activées-, tandis que celui de « donnée » renverrait plutôt soit aux formes de codification de la production sémiotique -p.e. données informatiques- soit aux stimuli physiques susceptible de produire l'information -données sensorielles-).

dimensions matérielles et communicationnelles de la transaction se confondent (cf. infra) ;

- ses *effets potentiels* psychiques et sociaux qui correspondent aux *conséquences possibles* de l'évocation de certaines représentations qui permettent d'attester de l'effectivité de la communication. Ces effets sont plus ou moins prévisibles selon le degré de standardisation de la situation transactionnelle et portent notamment sur l'actualisation du terrain commun et l'élargissement du « contenu sémiotique établi » entre les participants, considéré comme ayant publiquement ou officiellement fait progressé la transaction⁶ (les *effets potentiels* correspondent à une des facettes traditionnellement étudiée par la pragmatique linguistique).

On remarquera que ces définitions différencient au sein du média, le choix de la modalité d'expression et celui du support matériel associé à cette modalité d'expression, une distinction qui aura son importance pour l'analyse du document.

2.3 Distribution des transactions dans un cadre spatio-socio-temporel hétérogène

Du fait la distribution des activités collectives, réalisateurs et bénéficiaires des transactions communicationnelles peuvent ne pas être présents dans le même cadre spatio-temporel. Cela implique de donner une forme pérenne aux objets de la transaction de manière à ce qu'elle puisse être initialisée, interrompue, réactualisée, répétée, dans toutes les configurations de présence absence du bénéficiaire et du réalisateur. Par ailleurs, dans certains contextes, des réalisateurs distants peuvent se substituer aux réalisateurs ayant initié la transaction en « jouant leur rôle ». Dans ces contextes fréquents dans les organisations complexes et le monde économique nous parlerons de *distribution spatio-socio-temporelle des transactions*.

Dans la théorie socio-psycho-économique élargie des activités transactionnelles que nous sommes entrain d'élaborer, les média qui sont produits et échangés dans les transactions n'ont pas qu'une dimension sémiotique. Si les objets produits par les acteurs sont considérés comme des média, c'est avant tout parce qu'ils *médiatisent* les relations entre les acteurs en leur conférant un caractère transactionnel. Le fait qu'ils « véhiculent une signification » n'est qu'une de leur caractéristique majeure. L'autre dimension importante des média est leur dimension physique qui fait référence à des effets sur le bénéficiaire et qui reposent principalement sur la satisfaction de besoins d'ordre plus « corporels » ou « sensoriels » que « psychiques » (ce dernier terme regroupant des dimensions cognitives et affectives ou sentimentales).

Bien que les deux composantes soient toujours présentes, quand la dimension sémiotique prédomine sur la dimension physique du média, comme c'est le cas pour une production langagière qu'elle soit orale ou scripturale, nous parlerons de média sémiotique par opposition aux média physiques dans lesquels l'autre dimension prédomine. Les média symboliques ont aussi une dimension physique qui correspond aux effets marginaux exercés par leur *support* sur les organes sensoriels. La modification des supports autorisera aussi une extension de la fonction sémiotique qui sera, comme nous le verrons, exploitée dans le processus de documentarisation.

⁶ Correspondant à la notion de « Discourse Record » chez Clark.

Dans le contexte des transactions dont la production, principalement sémiotique, repose sur des médias symboliques (transactions communicationnelles), il existe différents moyens permettant de distribuer la transaction dans un contexte spatio-socio-temporel hétérogène. Nous en présenterons ici quatre, la *normalisation de la situation transactionnelle*, la *ritualisation mnémotechnique*, la *formalisation de l'expression* et la *documentarisation* pour insister sur ce dernier. Ces différents moyens ne sont pas complètement indépendants les uns des autres et se renforcent souvent. Dans le contexte de transactions impliquant principalement des *médias physiques*, d'autres stratégies comme l'automatisation, la formalisation des dimensions ou la conservation sont mises en place pour assurer la distribution des « transactions physiques ».

2.4 Quatre stratégies de distribution

La normalisation de la situation transactionnelle peut s'appliquer dans le contexte de la distribution de transactions communicationnelles ou physiques. Elle consiste à jouer sur les paramètres de la situation transactionnelle, homogénéisation des caractéristiques des acteurs en présence, reproduction des objectifs, des relations sociales, des compétences, des conditions environnementales, de manière à encourager la génération de productions sémiotiques similaires dans un cadre spatio-socio-temporel hétérogène. Cette normalisation n'est pas sans conséquences sur d'autres stratégies consistant à stabiliser le média lui-même en influençant, par exemple, les compétences requises pour la réalisation/réception de celui-ci et vice versa.

La ritualisation mnémotechnique consiste à exploiter les méthodes utilisées par les acteurs pour réaliser et recevoir la production sémiotique et en particulier pour la réactiver dans un contexte spatio-socio-temporel différent. Elle consiste à associer à la production sémiotique (à son média ou à son contenu) d'autres productions sémiotiques possédant avec elle des liens d'association de nature potentiellement très divers et qui sont susceptibles de multiplier les ancrages mnésiques dans le psychisme des acteurs, dans leurs automatismes corporels ou dans des caractéristiques stables de l'environnement matériel dans lequel les transactions sont susceptibles de se dérouler. Un accent particulier a été mis récemment sur ces stratégies dans le cadre de l'analyse du rôle de la corporéité dans le travail de mémoire (cf. par exemple Bachimont 2004). Des exemples très variés peuvent être fournis, allant des techniques de récitation aux techniques d'association mentale en passant par l'utilisation d'artefact externes stimulant la réalisation de la production sémiotique. Ses stratégies sont notamment utilisées par le même acteur engagé dans une transaction temporellement distribuée avec lui-même pour réactiver et en l'occurrence se remémorer une production sémiotique.

La formalisation de l'expression consiste à travailler sur cette composante du média de manière à stabiliser le contenu sémiotique. La formalisation de l'expression aura bien sûr des conséquences sur les conditions de réalisation et de réception de la production sémiotique. Mais à la différence de la ritualisation ou de la normalisation de la situation transactionnelle qui n'auront que des effets indirects sur celle-ci, elle vise préalablement à transformer l'expression du contenu sémiotique lui-même ce qui ne sera bien sûr pas sans conséquences sur ce contenu, sur son pouvoir évocation et ses effets potentiels. La formalisation de l'expression s'exerce dans les dimensions syntagmatique et paradigmatique. Dans le premier cas, elle vise à fixer précisément les conditions d'agencement des signes composant la production sémiotique par la détermination d'un ensemble de contraintes d'enchaînement. Dans le second cas, elle vise à fixer dans un réseau sémantique l'ensemble des signes

équivalents ou proches et la nature de leurs associations en fonction de leur pouvoir d'évocation et de leurs effets potentiels dans une situation transactionnelle donnée.

La documentarisation agit également sur le média de la production sémiotique. Mais au lieu d'influencer la modalité d'expression et sa formalisation, elle consiste à travailler sur le *support matériel*, ce qui comme on le verra, a aussi des conséquences importantes sur la modalité d'expression du contenu sémiotique. Dans la théorie des transactions communicationnelles, le *support matériel*, qui renvoie aux effets « physiques » du média, peut-être notamment classé selon qu'il est de nature éphémère ou pérenne. Les *supports éphémères* sont caractérisés par le fait que les stimuli qu'ils véhiculent ne peuvent exercer qu'un effet transitoire sur les organes des sens des récepteurs notamment du fait que ceux-ci ne peuvent exercer de contrôle direct sur le déclenchement et la reproduction des configurations de stimuli qui véhiculent les signes (c'est le cas des supports gestuels et vocaux produits par un interlocuteur dans le cas d'une interaction langagière en face à face, par exemple).

Par contraste, les *supports pérennes* sont susceptibles d'exercer un effet réitéré sur les organes des sens des récepteurs notamment grâce au fait que ceux-ci peuvent exercer un contrôle sur les conditions dans lesquels les stimuli sont rendus accessibles à leurs organes sensoriels dans des intervalles de temps courts (re-lecture de quelques lignes ou re-vision d'un extrait de film) ou longs (reprise d'un document après plusieurs jours, semaines, années...). Toutes les configurations intermédiaires sont envisageables, certains supports d'enregistrement, bien que « durables », n'offrant par exemple que des moyens restreints de manipulation des supports et donc de « navigation » entre les signes.

La stratégie de documentarisation consiste précisément à *transcrire ou à enregistrer les contenus sémiotiques sur des supports pérennes puis à doter ces supports d'attributs spécifiques permettant (i) de faciliter leur gestion parmi d'autres supports, (ii) de faciliter leur manipulation physique, condition d'une navigation sémantique à l'intérieur du contenu sémiotique et enfin, (iii) de faciliter l'orientation des récepteurs, mais également de plus en plus des réalisateurs eux-mêmes, à l'intérieur du support en définissant une ou plusieurs cartographies des contenus sémiotiques susceptibles de guider la navigation sémantique*. Parce qu'ils nous intéressent au premier plan dans le cadre de cet article nous allons évoquer les différents processus de documentarisation associés à cette stratégie.

3 Processus et moyens de la documentarisation

3.1 Transcription et enregistrement

Deux grandes méthodes permettent aujourd'hui de véhiculer une production sémiotique sur un support pérenne, la *transcription* et *l'enregistrement*. La première s'appuie sur la reproduction sur le support de signes graphiques dans des logiques de figuration ou d'équivalence phonétique, par exemple. Les signes graphiques codifiés les plus utilisés sont ceux qui sont utilisés dans les très nombreux systèmes *d'écriture* produisant des « textes ». Alors que la transcription implique le recours à un système de signes spécifiques et donc la transposition du contenu sémiotique selon une modalité d'expression nouvelle qui n'est pas sans conséquence sur son pouvoir d'évocation et ses effets potentiels, l'enregistrement peut sembler avoir des effets moins marqués sur les contenus sémiotiques faisant l'objet des transactions. Mais cette différence est peut-être moins profonde qu'il n'y paraît. En effet, si le

recours à la transcription possède un coût initial qui peut sembler plus important que celui de l'enregistrement, celui-ci présente le risque d'offrir une quantité de matériel considérable et difficile à exploiter et donc d'être plus difficile à documentariser notamment en ce qui concerne la cartographie des contenus sémiotiques véhiculés sur le média.

A la fois pour des raisons d'efficacité de la transaction et de documentarisation, l'enregistrement s'appuie souvent sur des stratégies de « mise en scène » des éléments de la situation transactionnelle qui induisent l'utilisation de techniques particulières de capture (associées dans certains cas à des techniques de montage a posteriori) définissant, comme dans le cas du recours à l'écrit, une modalité d'expression spécifique. Ainsi, de même que la transcription ne laisse pas le contenu sémiotique inchangé, l'enregistrement, qui pour être efficace doit être sélectif, transforme également ce contenu.

Enfin, de nouveaux supports hybrides et complexes ne cessent d'apparaître dans le contexte de la numérisation, dans laquelle le support ne sert plus seulement, par exemple, à reconstituer des textes en leur donnant une apparence aussi proche que possible de celle qu'ils pouvaient avoir sur le papier, ou à traduire sous la forme d'icônes un langage de commande, mais également à mettre en scène des situations transactionnelles dans le cadre d'interfaces de « réalité virtuelle » qui tout en pouvant donner l'impression de résulter d'un enregistrement sont en fait des constructions entièrement artificielles dans lesquelles de nouvelles modalités d'expression des contenus sémiotiques sont définies.

3.2 Définition du document et de la double articulation documentaire externe/interne

Mais on l'a vu, la transcription ou l'enregistrement sont des conditions nécessaires mais non suffisantes de la documentarisation. Une prise de note personnelle ou un enregistrement ponctuel visant à faciliter une activité sémiotique dans une situation transactionnelle donnée pourront avoir une utilité locale certaine. Cependant, s'ils ne font pas l'objet d'un *investissement documentaire* permettant de soutenir les processus de documentarisation, ces supports ne pourront pas être *ré-exploités* dans des situations transactionnelles différées et distribuées. En toute rigueur, il aurait été possible de différencier des stratégies de transcription et d'enregistrement des stratégies de documentarisation stricto sensu dans l'analyse des moyens permettant la distribution des transactions, mais celles-ci sont en fait très intriquées, transcription et enregistrement se faisant rarement sans une perspective même vague de ré-exploitation et donc de documentarisation.

Cette dernière considération nous permet de fournir notre définition du document en complément de celle de la stratégie de documentarisation. Le document désigne une *production sémiotique transcrite ou enregistrée sur un support pérenne qui est équipée d'attributs spécifiques visant à faciliter les pratiques liées à son exploitation ultérieure dans le cadre de la préservation de transactions communicationnelles distribuées*. Ces attributs doivent permettre au document de circuler à travers l'espace, le temps, les communautés d'interprétation, pour tenter de prolonger les transactions communicationnelles initiées par ses réalisateurs. Ils contribuent à une double articulation documentaire. Comme nous le verrons plus bas, la notion de production sémiotique tend à exclure, ou du moins à mettre à la marge du champ documentaire la production de données automatisées, les transactions associées à ces données pouvant être considérées comme des transactions plus matérielles que communicationnelles.

Pour progresser dans l'analyse des stratégies de documentarisation, il est nécessaire de rentrer dans la diversité des pratiques associées, d'une part, (i) à « l'exploitation externe » du document parmi d'autres documents avec lesquels il est stocké à l'intérieur de *bibliothèques, d'archives, d'armoire de classement, de base de données administratives, techniques ou documentaires* de nature diverses (de grande taille ou non, personnelles, privées ou publiques, gérant des supports numériques et/ou papiers, regroupant des productions sémiotiques plus ou moins hétérogènes dans leur forme d'expression ou dans leurs « genre », etc.) et d'autre part, (ii) à « l'exploitation interne » du document qui peut lui-même regrouper un ensemble de sous-productions sémiotiques dont le nombre peut être très variable et dont les modalités d'articulation peuvent également être fort diverses.

Pour répondre aux besoins du premier type d'exploitation du document, celui-ci sera doté d'un certain nombre d'attributs qui permettront son *articulation sémiotique externe* avec d'autres documents renvoyant à des projets transactionnels plus ou moins équivalents. Ces attributs permettront de dater le document, de le localiser, d'identifier les réalisateurs de la transaction (auteurs) ou les types de bénéficiaires auxquels il s'adresse (lecteurs), de décrire de manière schématique son contenu sémiotique, etc.

Pour répondre aux besoins du second type d'exploitation, on utilisera des attributs qui permettront de décomposer le document en parties cohérentes (titre, espacements, table des matières...), de mettre en relief certaines expressions permettant l'orientation sémiotique du lecteur à l'intérieur du contenu sémiotique (sous-titrage, typographie...), de renvoyer le bénéficiaire sur d'autres parties sémiotiquement liée (renvoi, index, notes...). Ces attributs qui relèvent de *l'articulation sémiotique interne* constituent un système d'orientation permettant au bénéficiaire de la production sémiotique, au lecteur dans le cas d'un texte, de naviguer sémiotiquement à l'intérieur du document.

3.3 Liaisons organique et mécanique entre les fragments d'un document

L'utilisation d'un support pérenne pour médiatiser la circulation des contenus sémiotiques va également permettre d'inscrire ou d'enregistrer sur le même support un nombre très important de signes et donc potentiellement de transactions communicationnelles distinctes. Derrière un projet transactionnel principal résumé par le titre du document, celui-ci regroupe un nombre important de sous projets transactionnels correspondant à autant de fragments documentaires plus ou moins étroitement reliés entre eux sans que la diversité de ces sous projets ne soit toujours explicitée. L'enjeu de l'orientation sémiotique du « lecteur » à l'intérieur d'un document renvoie donc à la question complexe de l'articulation interne des fragments dont le document est composé ou encore de leur mode de liaison.

Pour y répondre, nous établirons une distinction entre deux systèmes d'orientation en nous basant sur une analogie terminologique avec la sociologie qui établit une distinction entre une « solidarité organique », à la fois plus naturelle et tacite et « une solidarité mécanique » plus artificielle et formelle entre les membres d'une société. Les liaisons organiques sont celles qui correspondent à l'enchaînement naturel des fragments qui exploite les différentes logiques de planification associées au projet transactionnel du réalisateur : temporelle, structurelle, fonctionnelle, etc. Elles exploitent également toutes les relations de présupposition autorisées par le partage d'un terrain représentationnel commun, modifié et élargi au fur et à mesure du déroulement de la transaction. Ce découpage est la fois celui qui peut sembler le plus « naturel » et celui qui semble fournir une logique d'accès implicite au contenu sémiotique.

Le système d'orientation mécanique vient doubler le précédent en fournissant un système d'orientation supplémentaire qui soit, explicite le projet navigationnel initial (sous-titrage et table des matières), soit vient proposer des projets navigationnels a priori plus inattendus (index, tables de référence...). Dans ce dernier cas, il s'appuie le plus souvent sur un ensemble de ressources terminologiques et ontologiques organisées selon une logique différente de celle de la « table des matières », à la fois plus générale et plus systématique ou exploitant un point de vue très particulier, comme le nom des auteurs cités dans un texte à caractère littéraire. Sa constitution demande toujours un investissement significatif mais peut conférer au document une valeur ajoutée importante correspondant à une plus grande polyvalence transactionnelle.

L'exploitation de ces deux types de liaisons est le plus naturel dans le cadre de l'articulation interne. Mais il est également possible de retrouver ces logiques dans le cadre de l'articulation externe, notamment dans le cas de documents appartenant à des collections ou de petits documents regroupés dans des dossiers. Ainsi, le projet de collection pourra de manière implicite proposer une liaison organique entre les « volumes » qui pourra être doublée par un système de codification plus explicite ou par la mise à disposition d'un thésaurus global permettant une exploitation plus « transversale » de la production sémiotique.

3.4 Transactions codifiées et systèmes informatisés

Le recours au document a connu une extension considérable avec la généralisation de l'imprimerie puis avec celle de la numérisation qui correspond elle-même à une explosion des pratiques transactionnelles distribuées dans les univers de l'administration, de l'économie, de la politique, de l'art et du divertissement, de la science... Mais alors que certaines transactions sont longues, complexes, créatives, d'autres sont brèves, automatisées, standardisées. Dans la description des transactions intellectuelles (Zacklad 2000) une distinction est faite entre les *transactions automatisées ou matérielles*, dans lesquelles les dimensions sémiotiques et physiques du média se confondent et des *transactions investies de subjectivité* qui impliquent une activité interprétative permettant une évocation effective des représentations partagées et des engagements associés.

Les transactions du premier type qui correspondent, par exemple, à un acte d'achat automatisé, produisent des données très standardisées qui sont un cas limite de production sémiotique. C'est également le cas, par exemple, de l'échange monétaire stricto sensu dans la conclusion d'une transaction commerciale dont le contenu sémiotique se résume aux *effets potentiels* liés au transfert de biens dans le cadre d'un système conventionnel très précis. Pour désigner ces productions sémiotiques apparentées aux productions documentaires mais beaucoup plus pauvres en signification, nous parlerons de *production sémiotique codifiée* associées à des pseudo-documents, correspondant à des transactions automatisées, réalisées par des acteurs humains ou par des automates, dans lesquelles les dimensions sémiotiques et physiques du média se confondent.

Dans la littérature en informatique, il est souvent d'usage de désigner ces productions en parlant de « données structurées » gérées par les bases de données de gestion automatisées classiques, opposées aux données « semi-structurées » véhiculées par les NTIC et renvoyant aux documents numérisés eux-mêmes plus ou moins codifiés. Bien sur, les documents numérisés contiennent également des « données structurées » et il faut comprendre cette expression en considérant qu'elle sous-entend que dans les systèmes de gestion automatisés classiques, les données structurées tendent à épuiser la signification du pseudo-document

alors que la codification ne concerne que des zones restreintes dédiées à l'articulation mécanique dans le document semi-structuré. Mais en rentrant dans les détails, les choses sont toujours plus complexes, les systèmes de codification utilisés dans les transactions à base de données standardisées n'étant pas si univoques que leur concepteurs l'espèrent généralement.

Par contraste, la caractéristique des données semi-structurées véhiculées par les NTIC, qui correspondent à des documents numérisés au sens de notre définition, est généralement que leur interprétation ne peut pas se faire sans considérer les caractéristiques toujours pour partie variables de la situation transactionnelle qui n'est pas, dans la cas de l'usage des NTIC, complètement standardisée. De ce fait, la connaissance du nom du réalisateur de la production sémiotique, son auteur, est par exemple nécessaire à la bonne interprétation du contenu sémiotique. Ces informations devront donc être fournies dans le document pour faciliter son exploitation externe comme interne alors qu'elles ne le sont pas toujours dans le cas d'un formulaire de saisie.

Dans le cas des productions sémiotiques codifiées, le référencement à l'intérieur des bases de données est peu problématique dans la mesure ou du fait de la standardisation extrême de la situation transactionnelle, qui permet la formalisation de l'expression, il existe une isomorphie assez stricte entre le contenu sémiotique et les paramètres de la situation transactionnelle qui serviront à la recherche du pseudo-document. Dans le cas d'une facture simple, par exemple, l'intitulé, le montant, le destinataire, correspondent en principe exactement aux paramètres de la situation et aux effets de la transaction sur celle-ci. Dans la recherche de la facture, l'ensemble des paramètres pertinents de la situation transactionnelle, correspondant aux « champs du formulaire » sont prédéfinis et l'ensemble des contenu sémiotiques « échangeables », les valeurs possibles, restent dans des plages de variation elles aussi prédéfinies. Par contre, les conditions d'exploitation de ces données pouvant être assez diverses et nécessiter notamment des comparaisons avec d'autres informations, la problématique de la définition d'une « interface homme machine » peut être, elle, plus complexe à traiter.

Dans le cas des documents plus riches, le couplage entre les paramètres de la situation transactionnelle et la nature du contenu sémiotique est beaucoup moins étroit. De ce fait, la problématique du référencement du document par rapport à des « usages » assez ouverts est un problème en soi. La notion de « genre » du document, par exemple, utilisée en sciences de l'information (Aussenac & Condamines 2004), fait partie des efforts pour définir des classifications a priori des usages des documents pouvant être utilisées pour leur référencement. Elle se base sur une évaluation globale de l'intérêt du document pour une classe de lecteurs et d'une approche stéréotypée des situations transactionnelles associées. D'autres productions sémiotiques plus précises de type « mots-clefs » sont également élaborées pour faciliter la recherche, notamment dans la documentation technique. Elles se heurtent aussi à la difficulté d'anticipation des multiples usages possibles du document primaire.

3.5 Transactions documentaires

A cause de l'indétermination relative de certaines transactions communicationnelles qui peuvent viser des récepteurs nombreux aux intérêts variables, d'autres transactions spécifiques seront nécessaires à la recherche et au classement de ce document dans un système de rangement, correspondant à des « usage documentaires » spécifiques que nous différencions des « usages primaires » associées à la production sémiotique d'origine. Ainsi,

on ne confondra pas l'exploitation d'un document médical dans une situation de soin (usage ou *transaction primaire*), de la recherche de ce document à l'intérieur d'une bibliothèque (usage ou *transaction documentaire*). Les transactions documentaires peuvent également impliquer des acteurs différents de ceux associés aux transactions primaires.

Aux transactions sémiotiques de nature documentaire seront donc associées des productions sémiotiques spécifiques finalisées par la recherche et le classement du document. Dans le contexte de la documentarisation et en particulier de la professionnalisation associée à cette activité, ces productions sémiotiques seront elles-mêmes pour partie standardisées et inscrites sur une partie spécifique du support du document. Il s'agit notamment des ressources terminologiques et ontologiques, (RTO, pour reprendre l'expression d'Aussenac et Condamines 2004), également dénommées « méta-données », qui relèvent de la catégorie plus générale des *bases de codification* (cf. infra) incluant également différents systèmes de numérotation.

Les bases de codification incluant les ressources terminologiques et ontologiques et les systèmes de numérotation nécessitent, pour l'articulation documentaire externe, une standardisation des paramètres de la situation transactionnelle jugés pertinents comme le nom des réalisateurs, la date, le lieu de production etc. doublée de données relatives aux caractéristiques du support (nombre de page, localisation, etc.). Pour l'articulation sémantique interne, elles utiliseront notamment des ressources ontologiques, de nature soit formelle soit sémiotique (cf. infra), permettant la navigation dans les contenus sémiotiques.

4 Transaction communicationnelles émergentes dans les documents fragmentés : DopA et annotations

4.1 Documents lisses, documents fragmentés

Comme nous l'avons évoqué plus haut, un document, sous l'apparence d'un projet transactionnel unique décrit par son titre, n'atteint cet objectif que par le biais du recours à un nombre éventuellement très important d'arguments qui sont autant de sous transactions communicationnelle virtuelles et partielles participant de l'atteinte des objectifs du projet principal. Quand les documents ont une certaine longueur, c'est parce qu'ils fournissent un argumentaire détaillé dont les parties visent à répondre aux objections virtuelles qui pourraient être soulevées par les bénéficiaires auxquels il s'adressent. Dans une transaction différée, chaque ensemble d'arguments s'appuie sur un terrain représentationnel commun associé à des situations transactionnelles antérieures au sein d'une communauté, terrain commun qu'il sera souvent nécessaire d'explicitier et qui sera ensuite modifié par l'élargissement du contenu sémiotique établi au sein de cette communauté.

Dans le domaine des transactions à dominante expressives, qui poursuivent des projets de nature plus esthétique que normative ou intellectuelle (cf. Zacklad 2004), les transactions s'inscrivent dans une logique de récit descriptive et les configurations d'arguments sont alors plutôt liées à mise en place de différentes « scènes » au sein desquelles se déroulent les actions des personnages. Le succès des transactions littéraires de fiction tient sans doute en grande part aux conditions d'identification des lecteurs vis-à-vis transactions décrites qui dépendent de la vraisemblance et du plaisir associé à la découverte de nouvelles dimensions des expériences communes.

Plus le document apparaît comme « lisse », semblant écrit d'un seul tenant, plus les différentes prises en charges énonciatives correspondant à des sous-transactions communicationnelles sont implicites, les relations sémantiques entre celles-ci correspondant à une articulation organique dans laquelle les présupposés associés aux terrains représentationnels communs permettent un enchaînement naturel des « arguments ». Chaque technologie littéraire correspondant à un « genre » privilégie une modalité d'articulation organique spécifique de manière implicite.

Inversement d'autres documents apparaissent comme beaucoup plus fragmentés, les différentes sous-transactions communicationnelles étant mécaniquement articulées aux précédentes par le biais d'un système de numérotation, d'un sous-titrage, d'un ensemble d'attributs spécifiques visant à expliciter le statut des fragments dans le document et leur relation avec les fragments environnants. Cette fragmentation pourra être extrême dans une classe de documents jouant un rôle essentiel dans la coordination des transactions communicationnelles distribuées les « documents pour l'action ».

4.2 Production sémiotique coopérative : documents et dossiers pour l'action

Dans un nombre croissant d'activités professionnelles coopératives, notamment dans les activités à dominante « intellectuelles », le produit intermédiaire ou final de la coopération est une production sémiotique qui aura elle-même vocation à faire l'objet d'autres transactions à des échelles spatio-socio-temporelles plus grandes. Les collectifs engagés dans des activités coopératives étroites et structurellement ouvertes (Zacklad 2003a, 2003b) étant elles-mêmes de plus en plus distribuées spatio-temporellement et les environnements de travail de plus en plus numérisés et webisés, la stratégie de documentarisation apparaît comme la stratégie privilégiée. Celle-ci s'appuie sur différentes technologies de la coopération telles que, par exemple, les éditeurs de textes qui offrent des possibilités d'annotation tout en étant particulièrement bien intégrés aux outils de messagerie.

Mais les activités communicationnelles associées à la coopération structurellement ouverte soutenue par des supports pérennes génère des documents très différents des documents qui sont, par exemple, rangés dans des bibliothèques stockant des « œuvres » au sens traditionnel de ce terme. Ces documents, les « documents pour l'action » (DopA), sont pourtant de plus en plus indispensables aux activités collectives coopératives et ils correspondent à un nombre croissant d'usages collectifs des TIC dans le contexte des projets finalisés. Dans certains cas, quand ils regroupent des productions correspondant à des situations transactionnelles distinctes mais fortement interdépendantes, nous parlerons de « dossier pour l'action ».

Les exemples de DopA sont très divers et correspondent à de nombreux contextes professionnels. Citons, sans être exhaustif, les documents de conception en ingénierie (mécanique, logiciel...), les dossiers patients en médecine, les documents contractuels dans un contexte d'affaires, qui passent du stade de proposition commerciale au stade contrat en bonne et due forme, les dossiers qualités de plus en plus souvent numérisés, les rapports d'étude dans le conseil en management, les forum d'échange dans le domaine, par exemple, du logiciel libre...

Leurs caractéristiques principales sont les suivantes :

- Inachèvement prolongé : ils possèdent un statut d'inachèvement prolongé pendant la phase active de la production sémiotique collective phase durant laquelle nous les appellerons des *DopA en évolution* (vs *DopA stabilisés*) ;

- Pérennité : ils possèdent une certaine pérennité associée aux engagements des acteurs l'égard des contenus sémiotiques véhiculés et à la distribution des transactions d'où des enjeux liés à leur documentarisation et corollairement à leur stockage et à leur indexation ;
- Fragmentation : au moins pendant leur phase évolutive ils articulent plusieurs fragments plus ou moins liés sémantiquement (notamment des annotations) qui ne peuvent être fusionnés organiquement ou intégrés mécaniquement à l'ensemble du document (cf. infra) ;
- Rapport non trivial aux réalisateurs : les différentes parties du DopA sont le plus souvent produites par différents réalisateurs (plurivocité, pluri textualité...) qui peuvent avoir des statuts variables dans la situation transactionnelle et donc vis-à-vis de la production sémiotique (droits, engagement, responsabilités...) ;
- Rapport argumentatif non trivial aux autres parties du document : chaque partie possède une relation potentiellement complexe à ces autres parties : modalisation, incertitude, relation argumentative vis-à-vis des autres fragments...

Ces caractéristiques des DopA font qu'ils sont à même de justifier de nouveaux principes d'indexation et de classement encore peu pris en compte dans la profession documentaire mais qui répondraient aux besoins de micro-bibliothèques numériques en apportant un soutien à des projets en cours d'activité puis, à travers l'archivage des différentes versions des DopA, aux problématiques de « gestion des connaissances » aujourd'hui de plus en plus critiques. La finalité d'une indexation n'est pas la même selon qu'elle s'inscrit dans un processus évolutif et sert l'activité coopérative d'un collectif restreint où qu'elle vise des documents stabilisés mis à disposition d'une vaste communauté d'utilisateurs potentiels. En particulier, dans le cas des DopA évolutifs, le document passe par un nombre important de versions au travers desquelles son statut va changer de même que celui des fragments qui le compose.

Par ailleurs, ils relèvent de technologies littéraires et documentaires très diverses correspondant à des degrés variables de standardisation de la situation transactionnelle qui induit elle-même une standardisation des rôles joués par les participants et du format de leurs productions sémiotiques. Dans un ordre de structuration décroissant et sans être exhaustif, on pourra ainsi contraster (1) l'usage de formulaire, (2) le recours à des plans types (par exemple, des modèles systémiques en ingénierie, ou le plan conventionnel d'un contrat juridique), (3) des conventions d'écriture correspondant à un style stéréotypé de type descriptif ou argumentatif, par exemple, dans le contexte de la littérature scientifique, (4) les systèmes de question/réponse utilisés dans les forums et enfin (5) les documents narratifs, dans lesquels les efforts de style visent au contraire, à masquer le plan sous-jacent. Dans chacune de ces technologies littéraires⁷ et documentaires, les rôles respectifs de l'articulation sémantique et mécanique varient de même que le degré global d'intégration du document (un forum, par exemple, recourt plutôt à une articulation mécanique nous semble moins intégrée qu'un document technique).

4.3 Transactions communicationnelles émergentes dans les activités coopératives

Pour bien comprendre la structuration des DopA, il est nécessaire d'analyser la dynamique transactionnelle qu'ils autorisent. Or, si ces transactions sont sans équivalent réel dans les

⁷ La notion de « technologie littéraire » est ici utilisée dans un sens très large puisqu'un formulaire est considéré comme relevant d'un « style » particulier...

technologies de la coopération précèdent la numérisation intensive des supports, elles s'apparentent, au moins métaphoriquement, aux transactions communicationnelles qui prévalent dans les activités coopératives synchrones. Dans ces situations, qui correspondent par exemple aux réunions de travail, les différents participants peuvent proposer des éléments de solution, se contredire, argumenter, dans le cadre d'interactions polylogales soutenues ou non par l'utilisation d'un support d'inscription collectif, tel qu'un tableau.

Dans les interactions polylogales, les transactions communicationnelles relèvent de processus de type émergent. En effet, dans ces contextes, la situation transactionnelle est en mesure d'être « re-configurée » rapidement notamment en ce qui concerne la sélection des participants à la transaction en ratifiant explicitement ou non des participants qu'ils soient présents ou absents, concrets ou idéalisés (une personne collective, par exemple). De même, la capacité des acteurs à s'exprimer en « nous » peut être analysée comme la capacité à proposer rapidement la constitution d'un acteur collectif défini par un self commun aux contours plus ou moins précisément définis. De la même manière, le recours à des formulations implicites ou l'usage d'indexicaux pour faire référence à différentes composantes de l'environnement rend le contenu sémiotique potentiellement ambigu. De ce fait, une transaction sémiotique initiée par un réalisateur dans une situation d'interaction polylogale est sujette à de multiples interprétations qui correspondent à différentes offres de transactions virtuelles qui ne pourront être considérées comme engagées qu'en fonction des réactions des autres participants.

Dans les interactions polylogales en présentiel, les participants cherchent en permanence à conférer une cohérence aux transactions en s'appuyant sur leurs connaissances des composantes de la situation transactionnelle et sur leurs compétences communicationnelles (du type règle de pertinence de Grice, 1979) pour désambiguïser les transactions. Celles-ci sont d'autant plus flexibles et potentiellement efficaces que les participants se situent dans un cadre spatio-socio-temporellement homogène qui leur permet de corriger en continu les erreurs d'interprétation supposées. Dans le contexte de transactions distribuées par le biais de DopA, il sera nécessaire de fournir des mécanismes supportant des formes de transactions émergentes en partie analogues, suscitées par le dispositif technique, mais ne pouvant bénéficier pour leur réalisation des ressources offertes par l'interactivité en situation présenteielle. Les dispositifs d'annotation sont un des principaux mécanismes utilisés.

4.4 Importance des annotations dans les DopA

La principale difficulté à laquelle sont confrontés les acteurs qui co-produisent un DopA est le manque d'information sur les paramètres du contexte transactionnel associé à la proposition d'un fragment, correspondant lui même à une offre de transaction. Lorsqu'ils interviennent sur le DopA, ils le font souvent en déposant sur le support des *fragments libres*, des éléments de production sémiotique parcellaires transcrits ou enregistrés mais dont les relations organiques et mécaniques avec les autres composantes du contenu sémiotique documentarisé sont implicites et atténuées. Ces fragments seront progressivement exclus ou inclus du corps du DopA lors du processus de documentarisation impliquant une reprise et une articulation portant à la fois sur l'expression et le contenu sémiotique. Mais si les fragments libres ne sont pas renseignés, c'est-à-dire articulés, au fur et à mesure qu'ils sont déposés sur le support, ce travail de reprise est impossible et le DopA ne peut alors soutenir efficacement les transactions émergentes distribuées associées à l'activité coopérative.

La technologie permettant l'articulation d'un fragment libre avec les autres parties du document est l'annotation. De même que toute transcription ou enregistrement sur un support pérenne d'une production sémiotique ne suffit pas à constituer un document, le fragment libre ne constitue une annotation qu'à l'issue d'un processus de documentarisation. Ainsi nous définirons *l'annotation au sens fort* comme un *fragment de production sémiotique documentarisé, c'est-à-dire doté d'attributs spécifiques permettant son articulation mécanique avec les autres composantes du document*⁸. L'activité d'annotation, qui ne se confond pas avec le dépôt d'un fragment libre sur un support documentaire, requiert donc un effort spécifique qui permettra la ré-exploitation des fragments dans le cadre d'une activité individuelle ou collective distribuée en restituant les éléments du contexte transactionnel associés à leur production⁹.

Parmi les articulations caractéristiques de l'annotation figure *l'ancrage* sur une zone du support véhiculant un contenu sémiotique spécifique et faisant l'objet d'une transaction communicationnelle en cours. Le fait de désigner de manière plus ou moins explicite le réalisateur de l'annotation fait également partie du processus de documentarisation de même que toutes les modalités d'articulation sémantique mécaniques et externes traditionnellement utilisées (date, lieu...). L'utilisation de différentes conventions graphiques, de signes de ponctuation, de « méta-commentaires » dans le « texte » de l'annotation, participe également de l'articulation sémantique cette fois de manière plus interne et organique, donc implicite. Notons que si l'annotation peut contribuer au développement du contenu sémiotique, elle peut également ne viser la modalité d'expression à des niveaux qui peuvent être très variables : style, orthographe, typographie...

4.5 Codification de l'investissement documentaire dans les technologies de coopération

L'édition et l'exploitation semi-automatique ultérieure des annotations au sein d'un document numérique implique une *codification* des ses attributs, que ceux-ci concernent l'articulation externe du fragment au DopA ou son articulation interne à des élément du contenu sémiotique en substituant au caractère tacite des liaisons organiques une justification explicite. Comme on l'a vu dans l'exemple de la facture, la codification permet d'associer de manière univoque des composantes de la situation transactionnelle (le réalisateur, des éléments du cadre spatio-temporel, des caractéristiques du support, des éléments de contenu sémiotique...) à un contenu sémiotique descriptif standardisé inscrit sur une zone du support (un nom propre, un numéro, un « mot-clef »...). Pour réaliser cette association, la codification doit puiser dans un ensemble de ressources prédéfinies (une base de codification) ou résultant d'un processus de génération systématique correspondant à une fonction au sens mathématique.

L'élaboration de la base de ressource de codification, ou base de ressource d'indexation, de même que le travail de codification des transactions, ou de classification, correspond à un investissement documentaire spécifique nécessitant un effort plus ou moins important. La qualité de la base de ressources facilitera ou compliquera la classification par rapport à

⁸ Cette définition n'est pas en opposition avec la définition récente donnée de l'annotation par Bringay, Carry et Charlet (2004), qui analysent celle-ci comme la trace des représentations mentales évoquées au sujet de la cible. Nous sommes surtout plus exigeant dans notre définition : l'annotation au sens fort implique une forme de documentarisation pour soutenir les activités distribuées articulées sur le document.

⁹ Un fragment à vocation annotative qui ne serait pas ré exploitable par ses bénéficiaires potentiels à l'issue de sa création par défaut de documentarisation – ancrage peu clair, auteur incertain, signification ambiguë... - pourrait au mieux être considérée comme une *annotation au sens faible*.

l'utilisation d'une description ad hoc. Le retour de cet investissement est fourni lors de l'exploitation ultérieure du document, qu'il s'agisse de le retrouver ou de s'orienter à l'intérieur du contenu sémiotique qu'il véhicule, par la réduction du coût cognitif associé à cette exploitation. Dans un certain nombre de cas, les technologies de la coopération pourront tirer partie de la codification pour faciliter l'exploitation ultérieure du document, le travail coopératif du collectif, en facilitant les opérations de recherche et de filtrage. Parfois, elles permettront également d'automatiser partiellement une partie du travail de codification, réduisant de manière appréciable l'investissement documentaire des utilisateurs.

C'est le cas des systèmes d'annotation offerts dans les logiciels d'édition de texte qui offrent automatiquement une codification partielle de l'annotation selon un certain nombre d'attributs standards, auteur, date, ancrage, numérotation... Mais c'est également le cas d'autres technologies de coopération qui ne sont pas nécessairement appréhendées comme des technologies documentaires annotatives mais qui pourtant en possèdent plusieurs caractéristiques. C'est le cas de deux technologies dont les usages font l'objet d'études approfondies, le courrier électronique, basée sur une métaphore épistolaire et les forums de discussion, basées sur la métaphore de la participation à un système de dialogue à distance.

Dans le cas des forums, qui constituent une classe de DopA particulièrement importante, la pérennité du support et sa mise à disposition sur un serveur partagé permet à tous les contributeurs de rajouter des fragments de texte à un niveau quelconque de la « discussion » en cours, sans pouvoir toutefois choisir à l'intérieur du texte une zone d'ancrage précise. L'articulation sémantique se fait d'un point de vue externe par la codification de la date, de l'auteur, la reprise du titre de la contribution initiale dans la réponse, la localisation dans la hiérarchie des fils de discussions voire, dans certains dispositifs, par la sélection d'un type stéréotypé d'intervention (question, commentaire...).

Cependant, bien que les forums soient présentés comme des systèmes de dialogue à distance, des observations détaillées montrent qu'ils ne correspondent que faiblement aux modèles conversationnels traditionnels des polylogues en présentiel (Lewkowicz & Marcocchia, 2004). Comme souvent dans le cas des innovations importantes, les forums de discussion n'avaient aucun équivalent « réel » avant le développement des technologies documentaires numériques. Par ailleurs, la codification des transactions offerte par les forums est assez souvent insuffisante pour permettre une ré-exploitation aisée des contenus sémiotiques déposés et ce dans le contexte d'une forte hétérogénéité du cadre spatio-socio-temporel qui implique, notamment, une compréhension des transactions en cours par des acteurs n'étant pas les « auteurs » initiaux des contributions effectuées. En conclusion, nous présenterons des pistes de recherche relevant d'un programme de socio-informatique (associant CSCW et ingénierie des connaissances) qui permettrait d'améliorer les technologies de la coopération actuelles en s'appuyant sur une compréhension plus fine des mécanismes transactionnels en jeu.

5 Conclusion : nouveaux dispositifs pour la gestion des bases de codification, web socio sémantique et ontologies sémiotiques

Dans les recherches en CSCW et en Ingénierie des Connaissances la compréhension des processus cognitifs et sociaux est mise à profit pour proposer de nouvelles caractéristiques fonctionnelles pour les technologies de la coopération ce qui, en retour, grâce au double processus (i) de description systématique et détaillée associé à la spécification des nouveaux

scénarios et (ii) d'évaluation des usages des maquettes réalisées, permet de générer de nouvelles hypothèses de recherche visant à mieux comprendre le fonctionnement des transactions communicationnelles et leurs conditions de réalisation.

Dans les recherches de Lewkowicz et Marcoccia (2004), par exemple, on propose notamment un nouvel attribut pour l'articulation sémantique externe d'un fragment selon que les bénéficiaires de la transaction proposée sont un participant ratifié ou l'ensemble du collectif intervenant sur le forum. L'exploitation de ce nouvel attribut a des conséquences sur l'interface homme machine pour « l'intervention dans le forum » comme pour la visualisation des transactions antérieures. Dans les projets Médiapro et Médianotte (Zacklad et al. 2003), l'objectif est de proposer de nouveaux attributs pour des annotations dans les DopA qui tiennent compte de différentes dimensions ; la « micro-organisation » (auteur, statut, date, projet en cours...), les « connaissances du domaine » (les domaines professionnels dont relèvent l'annotation correspondant à une indexation thématique), « l'argumentation » (formulation de contrainte, suggestion, critique, évaluation...).

La définition au sein d'un collectif engagé dans un projet commun d'une base de codification commune fait l'objet d'un programme du laboratoire Tech-CICO correspondant au développement d'un Web-Socio-Sémantique (W2S). Complémentaire au projet du Web Sémantique, qui met principalement l'accent sur la définition « d'ontologies formelles » (universelles et totalement standardisées), le projet du W2S vise à fournir des bases de codification malléables et adaptées aux besoins de codification évolutifs des communautés d'utilisateurs vivantes. Au lieu d'ontologies formelles, le W2S propose de recourir à des *ontologies sémiotiques* qui mettent en relation de manière moins univoque la situation transactionnelle et le contenu sémiotique et dont « l'interprétation » s'appuie sur la nature des relations entretenues par un *thème* de l'ontologie avec les thèmes environnants dans une logique multi points de vue¹⁰. L'exploitation des DopA, intrinsèquement inachevés et confrontant les points de vue de communautés d'utilisateur diverses, requiert selon nous l'usage de ces nouvelles techniques de codification, moins « calculables » mais beaucoup plus proches des besoins et des usages des communautés utilisatrices.

6 Bibliographie

Aussenac, N., Condamines, A. (2004). Documents électroniques et constitution de ressources terminologiques ou ontologiques, *Information-Interaction-Intelligence*, Volume 4(1) : 75-94.

Bachimont, B. (2004). *Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*. Habilitation à diriger des recherches de l'Université Technologique de Compiègne.

Briet, S. (1951). *Qu'est-ce que la documentation?* Paris: EDIT.

¹⁰ La gestion de ces ontologies sémiotiques se fait à l'aide du standard « HyperTopic » lui même implémenté dans la plateforme Agorae (p.e. Cahier et al 2004). Celle-ci vise à permettre la gestion évolutive par diverses communautés d'utilisateurs des ontologies sémiotiques. Les différents points de vue sont regroupés selon différentes dimensions qui relèvent aujourd'hui de la définition : (1) des « terrains représentationnels communs » liées à l'activité, (2) de la définitions des acteurs individuels et collectifs participant aux transactions et (3) de la description des effets potentiels souhaités de la transaction (« acte de langage », modalisation, action sur le support, etc...).

- Bringay, S., Barry, C., Charlet, J., (2004). The Heath Record: Kernel of a Medical Memory, *Knowledge Management and Organisational Memory Workshop* (associated to ECAI'04).
- Brown, J. S., Duguid, P. (1995). The Social Life of Documents. In E. Dyson Ed., *The Social Life of Documents*. p. 1-18. New York: EDventure Holdings Inc.
- Buckland, M. K. (1997). What is a "document"? *Journal of the American Society for Information Science*, 48(9), p. 804-809.
- Buckland, M. K. (1998). What is a "digital document"? *Document Numérique* (Paris), 2(2) : 221-230. <http://www.sims.berkeley.edu/%7Ebuckland/digdoc.html>
- Cahier, J.-P., Zacklad M. (2002). M. Towards a Knowledge-Based Marketplace model (KBM) for cooperation between agents, in Blay-Fornarino, M. Pinna-Dery, A., Schmidt, K., Zaraté, P., *Cooperative System Design* : 226-238, IOS Press.
- Cahier, J.P., Zacklad, M., Monceaux, A. (2004) Une application du Web socio-sémantique à la définition d'un annuaire métier en ingénierie, *Actes des journées Ingénierie des Connaissances 2004*, Mai 2004, Lyon.
- Clark, H. H. (1996). *Using language*, Cambridge University Press.
- Ducrot O. (1984). *Le dire et le dit*, Editions de Minuit.
- Grice, P. (1979). *Logique et conversation*, Communications, 30 : 57-72.
- Lewkowicz, M., Marccocia, M. (2004). The participative framework as a design model for newsgroups: PartRoOM, in Darses, F., Dieng, R., Simone, C., Zacklad, M., *Cooperative Systems Design* : 243-257, IOS Press.
- Lewis, D. K. (1969). *Convention: a philosophical study*. Cambridge MA: Harvard University.
- Otlet, P. (1934 [1989]). *Traité de documentation*. Brussels: Editiones Mundaneum. Reprinted 1989, Liège: Centre de Lecture Publique de la Communauté Française.
- Pédauque, R. T. (2003). *Document : forme, signe et médium, les re-formulations du numérique*. (Working paper) http://archivesic.ccsd.cnrs.fr/sic_00000511.html
- Zacklad, M., (2000). La théorie des Transactions Intellectuelles : une approche gestionnaire et cognitive pour le traitement du COS, *Intellectica* 2000/1, 30 : 195-222.
- Zacklad, M. (2003a). Un cadre théorique pour guider la conception des collecticiels dans les situations de coopération structurellement ouvertes, in Bonardi, C., Georget, P., Roland-Levy, C., Roussiau, N. *Psychologie Sociale Appliquée, Economie, Médias et Nouvelles Technologies*, In Press, (Coll Psycho), Paris :135-164.
- Zacklad, M. (2003b). Communities of Action: a Cognitive and Social Approach to the Design of CSCW Systems, in *Proceedings of GROUP'2003*: 190-197, Sanibel Island, Florida, USA
- Zacklad, M. (2003c). *Annotation collective en conception: théorisation de la notion de « Document pour l'Action » et de sa lecture/écriture hypertextuelle collective au travers du*

processus annotatif, journée Connaissances et Documents du GDR I3, Octobre 2003, Paris, <http://www.irit.fr/GRACQ/ACTIVITES/SEMINAIRE2003/ResumesDocetConnaissances.htm>

Zacklad, M., Lewkowicz, M., Boujut, J.-F., Darses, F., Détienne, F. (2003). Formes et gestion des annotations numériques collectives en ingénierie collaborative, in R. Dieng-Kuntz, *Actes des 14èmes journées francophones d'Ingénierie des Connaissances*, Presses Universitaires de Grenoble : 207-224.

Zacklad, M. (2004). Transactions communicationnelles symboliques et communauté d'action: une approche de la création de valeur dans les processus coopératifs, à paraître in *Actes du colloque de Cerisy : Connaissance, Activité, Organisation*, P. Lorino et R. Teulier, Eds, Maspéro.