Les documents et les annotations du dossier patient hospitalier

Sandra Bringay*, Catherine Barry*, Jean Charlet**

*LaRIA - UPJV, CNRS-FRE 2733,
5 rue du Moulin Neuf 80000 Amiens
{bringay,barry}@laria.u-picardie.fr
**STIM - DPA/DSI/AP-HP, 91, bd Hôpital 75634 Paris Cedex 13,
jc@biomath.jussieu.fr

Résumé

Les professionnels de santé ne disposent toujours pas d'outils informatiques de gestion du dossier patient hospitalier leur permettant de reproduire toutes les pratiques qu'ils réalisent avec le dossier papier. Nous nous intéressons ici à une de ces pratiques : l'annotation. En s'inspirant de la démarche de [12], nous adoptons une vision originale sur les documents en les considérant munis de leurs annotations. Notre objectif est de montrer l'intérêt des annotations numériques dans une telle application.

Mots-clés *: documents, annotations, dossier patient.*

Astract

Professionals of health still don't have at their disposal tools for managing the patient record, allowing them to reproduce all the practices they realised with paper medical reports. We focus our attention on one of these practices: annotations. Related to the work of [12], we adopt an original point of view about the documents with their annotations. Our aim is to show the interest of numerical annotations in such an application.

Keywords *: documents, annotations, patient record.*

1 INTRODUCTION

Les professionnels de santé utilisent traditionnellement un ensemble de documents papier, le dossier patient, pour véhiculer des informations

médicales. Ce dossier montrant désormais ses limites, de nombreuses équipes ont travaillé sur son informatisation depuis les années 80. Elles se sont confrontées à divers problèmes (technologiques, de conception, d'utilisation, etc.) et sont par ailleurs passées de problématiques liées aux traitements des données médicales par des logiciels (pour faire des recherches épidémiologiques, pour le pilotage médico-économique, etc.), à une problématique de gestion de l'ensemble des informations produites autour des patients par des logiciels (pour améliorer les pratiques de soins). En 2003, le projet DocPatient a été lancé à l'université d'Amiens dans le but d'informatiser le dossier patient en milieu hospitalier selon une approche documentaire[1]. Il est mené en collaboration avec deux sites pilotes[2] et un partenaire industriel[3].

Le projet DocPatient fait suite au projet Hospitexte[4], dans lequel le document a été considéré comme l'objet central du système informatisant le dossier patient sous la forme d'un hypertexte. Au début du projet DocPatient, nous étions confrontés à la difficulté de concevoir une vision consensuelle sur la notion de document, partagée par tous les membres du projet, qui plus est issus de disciplines différentes. Nos réflexions ont été éclairées par le document [12]. L'auteur y définit la notion de document et notamment les évolutions liées au changement de support (du papier au numérique). Il aborde la notion de document selon trois dimensions, en le considérant comme une forme, un signe et un médium. Nous avons mené une recherche bibliographique sur la notion de document[5] et nous nous inscrivons complètement dans les trois définitions de [12] pour étudier des documents particuliers : les documents médicaux.

En parallèle, nous avons étudié l'existant. Nous nous sommes intéressés tout particulièrement au dossier patient existant sur un de nos deux sites pilotes, le dossier du service de Pédiatrie II. Ce service dispose d'un dossier patient papier et d'un dossier informatisé, centré sur les données et utilisé pour faire des statistiques. Nous avons réfléchi à la manière dont nous pouvions transformer le dossier papier et le dossier

---

[1] Ce projet, financé par le programme HTSC, regroupe une équipe pluridisciplinaire composée des sciences pour l'ingénieur (informatique) et des sciences humaines et sociales (droit, gestion, psychologie). Site du projet : http://www.laria.u-picardie.fr/EQUIPES/ic/htsc

[2] Le service de Pédiatrie II (de Réanimation Pédiatrique et de Néonatologie dirigé par le Docteur G. Krim) du CHU d'Amiens et le Pôle Cardio-vasculaire de la Région de Picardie.

[3] La société UNI-MEDICINE http://www.uni-medecine.com/

[4] Ce projet a été réalisé au DIAM et achevé en 1999 par la soutenance de thèse de V. Brunie [4] (collaboration de B. Bachimont et J. Charlet).

[5] Rapport de veille *Documents numériques* RV/03/TV3/02, 2003. http://www.laria.u-picardie.fr/EQUIPES/ic/htsc_old/documents/rapport/RV-03-TV2-02.pdf

informatisé en un seul dossier informatisé. Nous nous sommes rendus compte que les professionnels de santé annotent souvent les documents papier car ils sont conçus de telle sorte qu'ils ne peuvent y consigner l'ensemble des informations qu'ils voudraient. Nous nous sommes alors demandé s'il était possible de construire des documents numériques plus pertinents permettant la saisie de l'ensemble des informations médicales ou bien s'il fallait transposer la pratique d'annotation des documents papier au numérique. Pour cela, nous avons étudié les situations où les professionnels de santé rédigent ou consultent les annotations[6]. D'après nos lectures [7 ; 8 ; 15], nous avons défini une annotation ainsi :

> *Une **annotation** est une note particulière attachée à une **cible**. La cible peut être une collection de documents, un document, un segment de document (paragraphe, groupe de mots, mot, image ou partie d'image, etc.), une autre annotation. À une annotation correspond un contenu, matérialisé par une inscription, qui est une trace de la représentation mentale que l'annotateur se fait de la cible. Le contenu de l'annotation pourra être interprété à son tour par un autre lecteur. Nous appelons l'**ancre** ce qui lie l'annotation à la cible (un trait, un passage entouré, etc.).*

Dans cette publication nous nous posons trois questions :

1. Dans la section 2, nous nous demandons quelles sont les particularités des documents médicaux et comment peut-on les informatiser. Dans le projet Hospitexte et [12], les auteurs ont souligné l'importance, pour un système d'information, de disposer d'une représentation de la structure des documents pour les traiter. Dans le projet DocPatient, nous approfondissons ce point.
2. Dans la section 3, nous adoptons une vision originale sur les documents en les considérant munis de leurs annotations. En effet, les annotations interviennent dans le contrat de lecture unissant le(s) auteur(s) aux lecteurs selon les trois axes de [12]. Nous nous demandons comment la possibilité d'annoter modifie le comportement des auteurs et lecteurs et en quoi la pratique d'annotation est pertinente dans une application informatique gérant des dossiers patient.

[6] Rapport de veille *Annotations* RV/03/TV3/02, 2003. http://www.laria.u-picardie.fr/EQUIPES/ic/htsc_old/documents/rapport/RV-03-TV2-02.pdf

3. En conclusion, nous adoptons un point de vue épistémologique et nous nous interrogeons sur les conséquences à tirer de l'étude de la définition des documents, sur leur statut, une fois annotés[7].

2 LES DOCUMENTS MEDICAUX COMPOSANT LE DOSSIER PATIENT

Nous adoptons les définitions de [12] pour considérer des documents particuliers : les documents médicaux. Nous nous demandons pour finir à quoi va ressembler le dossier patient informatisé.

2.1 Quels documents médicaux ?

La plupart des documents composant le dossier patient papier et informatisé du service de Pédiatrie sont des formulaires ayant évolués au cours des dernières années. Les formulaires sont des documents conçus par un concepteur pour être complétés par un rédacteur. Par exemple, le formulaire d'entrée du patient est un document créé par le chef de service pour être rempli par un infirmier. Il existe dans le dossier patient papier des documents qui ne sont pas des formulaires. Un chirurgien expliquant une opération à un patient, construit un schéma sur une feuille volante. Ce document n'a pas été conçu par un concepteur. En quelque sorte, il « improvise » le document en fonction de l'information à communiquer.

Dans cette section, nous détaillons les deux principales causes de l'évolution des documents médicaux, la place des formulaires en général et leurs interactions avec le concepteur, le rédacteur et le lecteur, ainsi que l'usage que les professionnels de santé en font. Nous voyons ensuite les notions de structure de documents et de documents (non-semi)-structurés. Pour finir, nous décrivons les documents médicaux que nous voulons utiliser dans le dossier patient informatisé.

[7] Dans cette publication, nous travaillons sur l'usage possible des annotations dans le cadre d'une application métier, le dossier patient, plutôt que sur la modélisation de cet objet complexe et son implantation informatique qui sont des travaux en cours. Sur ces thèmes, nous invitons le lecteur a se reporter aux travaux réalisés pour indexer des documents par l'ajout d'annotations sémantiques marquant le contenu [10 ; 13] ou pour communiquer des informations grâce aux annotations en relation avec des documents pour l'action [15].

Evolution des documents médicaux

Les professionnels de santé ont de plus en plus peur des conséquences médico-légales de leurs actes. En effet, les patients et la justice peuvent demander l'accès au dossier patient. D'après [14], les documents médicaux ne contiennent plus que les données essentielles obligatoires qui, d'après la loi, doivent être transmises aux patients. Les chefs de service imposent aux professionnels de santé d'utiliser des formulaires plus structurés et en partie figés, afin d'uniformiser les saisies. Ils fixent ce qui doit ou non figurer dans les segments des documents.

Par ailleurs, il y a eu une certaine évolution des mentalités. Dans le Service de Pédiatrie II, il y a plus de dix ans que les professionnels de santé utilisent une application informatique pour faire des statistiques. Ils se sont habitués à saisir les mêmes données dans les formulaires papier et numériques. Ils ont pris conscience de l'intérêt d'un tel outil pour la recherche et la communication d'informations, malgré le temps perdu en saisie. Ils préfèrent généralement utiliser le plus possible d'aides (cases à cocher, listes déroulantes) et taper le minimum de texte pour gagner du temps auprès du patient. Ils ont fait évoluer les formulaires sur format papier pour les rendre homologues aux documents numériques.

Une grande majorité de formulaires

Les instances hospitalières (les chefs de service) ont donc élaboré et imposé des formulaires types aux professionnels de santé. Nous détaillons ici les caractéristiques de ces documents et précisons les rôles des trois acteurs à l'origine de la conception et de la lecture d'un formulaire : le concepteur, le rédacteur et le lecteur.

Créer un document est une façon de traduire sa compréhension sociale, de gloser la réalité en communiquant, explicitant ou commentant des évènements. Dans le dossier patient, il s'agit d'événements ayant lieu lors du processus de soins. Le concepteur formalise le domaine « métier » en définissant l'organisation et le contenu du document. Par exemple, le chef de service concevant un compte rendu d'hospitalisation, s'attend à ce que le médecin inscrive certaines données dans les paragraphes du formulaire. Cette intention transparaît alors dans les intitulés descriptifs des champs. Grâce à son expérience de la pratique médicale, il définit le contenu et la structure du document en se fondant sur ses connaissances et savoir-faire métier. Le rédacteur perçoit l'organisation du document. Il répond aux

« questions » du concepteur en consignant les informations qu'il pense devoir y faire figurer suite à son interprétation des intitulés. Par la suite, un lecteur qui consultera ce document interprétera à son tour les intitulés du concepteur et les informations ajoutées par le rédacteur. « Un médecin qui lit un compte rendu d'hospitalisation sait d'avance que c'est un médecin qui l'a rédigé, pour l'un de ses pairs, qu'il s'agit d'un résumé des événements médicaux (…) survenus. Ce sont autant de prescriptions interprétatives lui indiquant comment lire le contenu. (…) Les genres textuels permettent donc la lecture d'un contenu dans de nouveaux contextes, distants dans le temps et l'espace du contexte de création » [3]. Les intitulés descriptifs sont autant de signes familiers pour le concepteur (qui les choisit), pour le rédacteur (qui saisit des informations) et pour le lecteur (qui interprète le document et les données qu'il contient). La lecture n'est possible que si le lecteur reconnaît le format documentaire pour *contextualiser* l'auteur, le lecteur visé et la relation qui les unit.

Dans le dossier patient papier, nous avons rencontré principalement deux types de formulaires :

1. Les formulaires rédigés en temps réel, au pied du lit du patient (p. ex. l'examen des fonctions respiratoires) : ces documents contiennent essentiellement des données brutes. Les professionnels de santé, qui ont en face d'eux les patients, ne veulent pas perdre de temps en saisie. Celle-ci se limite alors à quelques mots et segments de phrases.
2. Les formulaires rédigés *a posteriori* de l'acte médical (p. ex. le compte rendu d'hospitalisation) : ces documents contiennent des interprétations des données brutes. Ils sont conçus pour que le rédacteur dispose d'une certaine liberté de saisie car le concepteur ne peut prévoir l'ensemble des informations à conserver dans le document. Les titres des paragraphes servent alors de trames aux rédacteurs qui expriment les informations médicales en langue naturelle.

Il existe bien sur des documents mixtes composés de parties appartenant aux deux premières catégories identifiées.

Les différents niveaux de structure d'un document

Intéressons nous maintenant à la structure des documents. Dans le projet DocPatient, nous réalisons des calculs sur les documents et leur contenu en utilisant leur structure. Prenons quelques exemples. Pour faciliter la navigation de l'utilisateur dans le dossier, nous construisons des sommaires qui aident l'utilisateur à se forger une vision globale du dossier informatisé, en récupérant tous les titres d'un (ou plusieurs)

document(s). On exploite ainsi la structure de chaque document. Nous créons aussi des documents favorisant l'appropriation directe par un lecteur, des données du dossier informatisé. Les infirmières notent quotidiennement sur la pancarte le poids du nourrisson. Or, la pancarte correspond à une semaine d'observation et si un médecin veut connaître l'évolution du poids sur une plus longue durée, il doit consulter plusieurs pancartes. Nous créons donc un nouveau mode de représentation des données, sous forme d'un graphe, facilitant la perception humaine immédiate. On exploite alors la structure des données contenues dans les documents.

Nous nous plaçons donc ici dans l'équation de [12] :

Document numérique = structure + données

Selon [12], nous pouvons « décoder » le document en étudiant sa structure qu'il considère comme représentative du contrat de lecture liant l'auteur au lecteur potentiel. En effet, lorsque l'on considère un document, on se rend compte qu'il ne s'agit pas d'une entité formée d'un seul bloc mais, qu'au contraire, il comporte plusieurs segments (parties) distincts. Chaque segment contient plusieurs briques informationnelles (des données[8] et des informations) qui forment une certaine unité informationnelle, qui sont sources de signification. Le regroupement des segments en unités informationnelles plus importantes (p. ex. un chapitre dans un livre) a aussi un sens, fonction de l'intention de communication de l'auteur. Le lecteur doit posséder les connaissances nécessaires et un référentiel commun avec l'auteur pour s'approprier la structuration du document. On ne peut donc pas réduire un document en un ensemble de segments signifiants car leur organisation, l'agencement des segments entre eux (un segment comme titre introduisant des sous-titres et des paragraphes) et leur présentation (des mises en forme différentes entre les titres et les paragraphes), fait du document un tout indivisible.

Décrire la structure d'un document consiste à identifier et décrire chacun des segments le composant. Pour cela, nous distinguons le contenu du document, c'est-à-dire les briques informationnelles (données et informations) et les segments qui sont leurs conteneurs. Suivant le type de documents et l'application, le niveau de granularité entre ces deux niveaux varie. On différencie trois niveaux de description de la structure :

[8] Par abus de langage, nous utilisons le terme « donnée » pour désigner l'ensemble des données et des informations contenues dans les segments (les conteneurs). On ne parle pas ici du terme « donnée » utilisé en base de données pour désigner une information typée.

1. ***Décrire la structure physique d'un document*** consiste à décrire sa présentation. Si le document contient du texte, la description porte sur ses caractéristiques typographiques : police, couleur, taille, graisse.
2. ***Décrire la structure logique d'un document*** consiste à décrire le rôle et la nature de chaque segment, ainsi que l'ensemble des liens hiérarchiques et/ou logiques qui les lient les uns aux autres ***par rapport à l'ensemble du document.*** Pour décrire cette structure, on utilise des notions issues de l'édition : titre, sous-titre, paragraphe, etc.
3. ***Décrire la structure des données contenues dans les segments d'un document***, consiste à décrire l'organisation des données, leur rôle et leur nature, ainsi que les relations hiérarchiques et/ou logiques qui les lient ***dans un segment***. Pour décrire cette structure, on utilise des notions issues de la description du monde qui nous entoure : une coordonnée est composée d'une adresse et d'un numéro de téléphone, une adresse est composée d'une rue, etc.

Deux documents peuvent présenter des structures physiques différentes, prendre des formes perceptibles différentes, pour une même structure logique. Les structures logique et physique sont pourtant liées car cette dernière met en valeur la structure logique et facilite la lecture. Les documents papier du Service de Pédiatrie sont soumis à des conventions typographiques. Grâce à celles-ci, le lecteur se construit une image mentale de la structure logique du document et le reconnaît facilement, uniquement par son apparence. La structure logique d'un document est donc le fondement dont découle la structure physique. À l'inverse, la structure physique est une trace de la structure logique.

Ayant distingué les différents niveaux de structure du document, nous possédons désormais le vocabulaire nécessaire pour définir les notions de documents non structurés, semi-structurés et structurés.

Documents structurés, semi-structurés et non structurés

Lorsqu'un lecteur perçoit un document, s'il peut décrire les différents niveaux de structure, on distingue trois types de documents :

- ***Un document perçu comme non structuré*** est un document pour lequel le lecteur n'est capable de décrire ni la structure logique du document et ni la structure logique des données dans les segments.
- ***Un document perçu comme semi-structuré*** (ou « document orienté paragraphes » [9]) est un document pour lequel le lecteur est capable de décrire la structure logique du document. Par contre, il ne peut décrire la structure logique des données contenues dans les segments.

- ***Un document perçu comme structuré*** (ou « document orienté données » [9]) est un document pour lequel le lecteur est capable de décrire la structure logique du document (au moins un segment) et celle des données dans les segments.

Intéressons nous maintenant aux documents numériques qu'un lecteur visualise sur un support d'appropriation[9]. Le lecteur peut percevoir un document semi-structuré ou structuré sur son écran mais, la machine ne dispose pas forcément d'une représentation explicite de la structure logique de ce document. Par exemple, l'auteur d'une page Web, peut très bien structurer son document en panneaux. La machine ne dispose d'aucune information sur la signification de ce découpage logique. Par contre, connaissant la structure physique du document (les coordonnées physiques des panneaux), elle affiche le document conformément aux attentes de l'auteur. Au contraire, un document peut être structuré (le système dispose d'une représentation logique du document), sans que le lecteur perçoive cette structuration. Par exemple, une lettre, adressée aux familles, est générée automatiquement à partir des données contenues dans une base de données, qui viennent compléter des paragraphes en langue naturelle. Le lecteur visualise ces paragraphes textuels sans percevoir la représentation dont dispose le système. De même, un lecteur peut percevoir la structure d'un document sans avoir les connaissances pour l'interpréter. Par exemple, un patient examine un compte rendu d'opération structuré en paragraphes (hypothèses, conclusions, etc.) sans sous-titre. Ne disposant pas de connaissances médicales suffisantes, le patient ne pourra interpréter ce découpage.

Si le système informatique[10] possède des informations sur la structuration logique du document et des données dans les segments, on distingue trois types de documents numériques :

- ***Un document numérique non structuré*** est un document pour lequel le système informatique ne possède pas de représentation explicite de la structure logique du document et des données dans les segments.
- ***Un document numérique semi-structuré*** est un document pour lequel le système informatique possède une représentation explicite de la structure logique des segments mais ne possède pas de représentation explicite de la structure logique des données dans les segments.

---

[9] Comme [3 ; 12], nous distinguons le support d'inscription (ou support numérique virtuel d'opération) qui interprète informatiquement les enregistrements informatiques via les calculs (p. ex. le cédérom support d'un document) et le support d'appropriation (p. ex. l'écran) qui est interprété par un humain grâce aux formes sémiotiques matérielles calculées d'appropriation [2] (les pixels qui s'assemblent pour former des signes).

[10] On considère comme système informatique, la machine et son environnement.

- ***Un document numérique structuré*** est un document pour lequel le système informatique possède une représentation explicite de la structure logique des segments et des données dans les segments.

Conclusions sur les documents numériques médicaux

Nous voulons transposer les documents médicaux du support papier au support électronique sans chercher à modifier les structures physique et logique des documents et des données, actuellement connues par les professionnels de santé utilisant le dossier patient.

Le contrat de lecture entre le lecteur et l'auteur est fonction de la connaissance de ces différents niveaux de structure. Peu importe à l'auteur et au lecteur de savoir comment le document est stocké, si le système possède ou non une représentation explicite des différents niveaux de structure. Ce qui les intéresse, c'est la structure qu'ils perçoivent sur le support d'appropriation (p. ex. l'écran) et qui permet à l'auteur de transmettre un message interprétable par le lecteur. Posséder des informations sur la structure des documents et des données permet de créer de nouveaux documents. Nous pouvons, comme dans le projet Hospitexte, les traduire sous diverses formes perceptibles lisibles à l'écran, en occultant des éléments signifiants et en faisant redécouvrir de nouveaux éléments (création de documents de navigation, de synthèse).

En ce qui concerne les formulaires complétés en temps réel au pied du lit du patient, la rapidité de saisie et l'exploitation des données étant les critères essentiels, leur informatisation se fera sous la forme de formulaires structurés. En ce qui concerne les formulaires complétés *a posteriori* de l'acte médical, leur informatisation se fera sous la forme de formulaires semi-structurés pour laisser une certaine liberté de saisie aux rédacteurs, nécessaire à l'expression de leur analyse. Dans ces formulaires, les données présentes dans les paragraphes ne peuvent être traitées directement par un système d'information sans l'utilisation d'algorithmes de TALM (Traitement Automatique de la Langue Médicale comme dans le projet DRUID[11]). Il y a donc une perte d'information pour le système d'information et non pour le lecteur qui est capable d'identifier, au cours de sa lecture, les informations dans les paragraphes.

[11] Le projet DRUID [9] est réalisé au LIRIS (Laboratoire d'InfoRmatique en Images et Systèmes d'information) par entre autres A. Flory, D. Laforest, Y. Badr. Il a été intégré au projet DocPatient courant 2003. L'objectif est d'extraire des données médicales dans les segments de documents semi-structurés.

2.2 Quel dossier patient informatisé ?

Dans la partie précédente, nous avons précisé quels documents médicaux nous voulions voir apparaître dans le dossier patient informatisé. Nous nous demandons maintenant comment regrouper ces documents pour leur donner du sens et faciliter les pratiques de saisie et de lecture des professionnels de santé.

Le dossier patient papier est un hyperdocument[12] particulier, c'est-à-dire une collection de documents, faisant sens ensemble (car relatifs au même patient) et dont on peut préciser les liens qui les unissent (p. ex. le compte-rendu x est un document qui explicite la radio y). Pour évoquer cette unité des documents, [2] parle de *solidarité des documents*. Dans un tel dossier, il n'y a pas de parcours canonique de la lecture. Suivant son objectif, le professionnel de santé peut survoler un ensemble d'examens ou bien chercher une synthèse et la lire attentivement, etc. La multiplicité des parcours de lecture dans l'hyperdocument permet une lecture riche car suivant l'orientation du parcours choisie, le lecteur élaborera des interprétations différentes des mêmes documents.

Comme dans le projet Hospitexte, nous informatisons le dossier patient sous la forme d'un hypertexte[13]. Sous cette forme, la navigation est souvent problématique : la lecture n'est pas linéaire et le lecteur, rapidement désorienté dans la masse de documents, en oublie parfois son objectif et sa stratégie de lecture. En effet, un document cible influe sur l'interprétation du document source. [2] affirme même que la textualité du document cible remplace alors celle du document source, ce qui peut modifier l'objectif ou la stratégie de lecture. En effet, contrairement à un livre où tous les documents participent à la compréhension globale, dans un hypertexte, le lecteur n'est pas obligé de consulter tous les documents pour l'élaboration du sens. [2] parle alors de *solidarité annotationnelle* des documents de l'hypertexte, qu'il oppose à la *solidarité textuelle* des documents d'un livre. Dans un hypertexte, « tout document est l'annotation d'un autre » [2]. Nous pouvons chercher à compenser la perte de sens provoquée par la navigation en utilisant la *calculabilité* [2] du support numérique qui permet de créer des nouveaux documents, comme des sommaires, pour faciliter la lecture de l'hypertexte, en donnant un parcours plus ou moins canonique.

---

[12] « L'hyperdocument peut se définir comme étant la réunion d'un ensemble de documents constituant un tout, une unité, par rapport à une utilisation pratique déterminée. » [2]

[13] Selon [2], un hypertexte est un hyperdocument informatisé sous la forme d'un réseau de nœuds documentaires et de liens reliant ces nœuds les uns au autres.

3 LES DOCUMENTS ET LEURS ANNOTATIONS

Maintenant que nous avons précisé la manière dont nous comptions informatiser le dossier patient, nous nous intéressons aux documents et à leurs annotations. Comme [12], nous étudions le contrat de lecture sous trois dimensions : la lisibilité, la compréhension et la sociabilité et nous regardons en quoi les annotations interviennent dans ce contrat et quelles sont les conséquences pour la notion de document. Afin de montrer combien elles sont importantes pour notre application, nous prenons des exemples d'annotations rencontrées dans le dossier patient papier pour illustrer nos points de vue. Dans la dernière section, nous justifions le fait de considérer les annotations comme des documents.

3.1 Contrat de lecture

Lisibilité

Considérant le document comme une forme, [12] donne la définition :

« Un document numérique est un ensemble de données organisées selon une structure stable associées à des règles de mise en forme permettant un lisibilité partagée entre son concepteur et ses lecteurs. »

Dans les formulaires numériques, ce concept de lisibilité partagée se traduit par l'ajout du concepteur, de règles (les intitulés descriptifs) de saisies destinées au rédacteur et de lecture destinées au lecteur (*cf.* § 2.1.2). Or, l'ajout de ces règles ne permet pas de rendre compte de toutes les situations de saisie. Par exemple, les formulaires médicaux rédigés en temps réel, souvent très structurés, ne donnent pas assez de liberté aux professionnels de santé pour laisser des traces du contexte de rédaction. Les concepteurs ne pouvant prévoir l'ensemble des faits médicaux pertinents susceptibles d'être saisis, ajoutent alors des champs textes : les mémos. Les informations qui y sont consignées sont en général diluées dans du texte. Comme pour les paragraphes des documents semi-structurés, elles ne peuvent être traitées sans TALM. Une fois encore, il y a une perte d'information pour le système et non pour le lecteur, qui peut visualiser le contenu des mémos et identifier des informations dans ces

mémos. Toutefois, même pour le lecteur, ces mémos ne préservent pas le contexte de création de l'information ajoutée car ils ne la mettent pas en relation avec la partie du document ayant suscité le commentaire. C'est pour cette raison que sur les documents papier, les praticiens utilisent plutôt des annotations reliées à la partie annotée par un moyen graphique (flèche, surligné), pour ne pas perdre le contexte de création. Ainsi, les annotations ajoutées aux documents médicaux participent à la lisibilité des informations consignées dans le dossier, en conservant des traces du contexte de leur rédaction, que l'on identifie grâce à l'ancre (*cf.* § I).

Compréhension

Considérant le document comme un signe, [12] donne la définition :

> *« Un document est un texte dont les éléments sont potentiellement analysables par un système de connaissance en vue de son exploitation par un lecteur compétent. »*

Pour que le système ou un lecteur exploite les éléments significatifs contenus dans un document, ceux-ci doivent posséder les connaissances nécessaires pour interpréter le document et les éléments qu'il contient. Comment interprètent-ils alors les annotations attachées au document ?

Prenons l'exemple des annotations produites lors de la lecture. Le lecteur laisse des traces dans les gloses[14] (annotations), selon son objectif de lecture, qui sont significatives de sa lecture des documents. Par exemple, un médecin monte un cours sur les traitements *x*. Il veut trouver des exemples réels d'utilisation de ce traitement pour illustrer son cours. Il cherche des dossiers où ce traitement a été utilisé et il les annote. En effet, les formulaires du dossier patient sont conçus pour être rédigés et lus dans un certain contexte. Or, si le médecin lit un document dans un autre contexte, le concepteur n'a pas prévu de champ pour ces nouveaux commentaires. Un chef de service ne peut pas répertorier tous les documents utiles à l'enseignement et il ne va pas laisser dans chacun un paragraphe dédié rarement renseigné. Les annotations permettent alors l'ajout d'informations relatives à une vision originale sur le document.

La possibilité d'annoter les documents influe sur leur interprétation. En effet, annoter facilite la compréhension du document car le fait de paraphraser, d'ajouter des informations ou de synthétiser participe à la

[14] D'après le petit Robert, une glose est une annotation entre les lignes ou en marge d'un texte pour expliquer un mot difficile, éclaircir un passage obscur.

conception par l'annotateur d'une connaissance, durant le processus d'intériorisation de l'information consultée dans le document. Nous nous rapprochons ici du concept de *solidarité annotationnelle* de [2] (*cf.* 2.2) pour désigner le fait que la compréhension d'un document dans un hypertexte influe sur celle des autres documents rencontrés durant le parcours lectorial. Ce phénomène est d'autant plus vrai lorsque le lecteur devient l'auteur d'un nouveau document : l'annotation. D'ailleurs, *a posteriori,* les annotations sont en général plus faciles à relire par l'annotateur que le document car formulées avec ses propres mots. Toutefois, lorsque les actes d'annotation et de consultation des annotations sont éloignés dans le temps, l'annotateur a parfois du mal à se comprendre. On oublie souvent quel était son objectif de lecture lors de l'annotation. Or, quand on lit un document, si l'objectif de lecture est différent, on l'interprète autrement et les annotations sont différentes.

Á partir de ces annotations, il est possible de construire de nouveaux documents (« documents de lecture » [5]), qui correspondent aux points de vue retenus par le lecteur sur un document. Par exemple, un professionnel de santé, parcourt le dossier patient à la recherche des causes possibles d'une complication et surligne certaines informations. Regrouper dans un nouveau document tous les segments de phrases surlignés a un sens. Ce genre d'application est possible dans des cadres applicatifs particuliers (une suite d'annotations extraites de leurs contextes originels, les segments de documents annotés ou le document lui-même, est rarement exploitable) et sur des annotations particulières, susceptibles de prendre progressivement leur indépendance vis-à-vis du document annoté, pour devenir des documents elles-mêmes. Prenons l'exemple d'un *post-it* collé sur la pancarte « Penser à changer la perfusion x ». Une telle annotation ne sera jamais autonome car elle fait forcément référence à un autre objet, la perfusion elle même. Au contraire, une annotation comme la glose (p. ex. le commentaire critique d'un examen) peut acquérir un certain degré d'autonomie et prendre tout son sens dans le contexte d'un dossier. Une telle annotation est alors compréhensible, au moins en partie, seule[15].

Les annotations sont donc une bonne solution pour donner aux praticiens les moyens de construire leur propre lecture des documents et de générer leur propre vision du dossier. « En y apportant sa propre marque, le lecteur devient l'auteur de sa propre lecture dans la mesure où il reformule dans ses propres termes ce qu'il voit : en quelque sorte, il écrit le document qu'il aurait voulu consulter s'il avait été l'auteur. » [1]

[15] Cette question de l'autonomie des annotations mérite encore réflexion (*cf.* conclusion).

Sociabilité

Considérant le document comme un médium, [12] donne la définition :

« Un document numérique est la trace de relations sociales reconstruites par les dispositifs informatiques ».

Pour [12], le document établit un lien social entre différents acteurs (auteurs et lecteurs) et acquiert par ce biais un statut social.

La réorganisation du temps de travail dans les services hospitaliers a accentué le problème de la transmission d'informations entre les différentes équipes se relayant autour des patients. Dans le dossier papier, on trouve des annotations qui servent de support à ces échanges d'informations. Par exemple, lorsque les professionnels de santé rédigent un document, ils ajoutent souvent des commentaires dans des *post-it* ou dans la marge, s'adressant aux collègues leur succédant. Ils communiquent de manière asynchrone [6] (en temps différé) dans le contexte d'un document.

Avec le numérique, nous pouvons aussi imaginer faire communiquer de manière synchrone [6] (en temps réel), deux professionnels de santé dans le contexte d'un document, toujours grâce aux annotations. Par exemple, un professionnel de santé demande à un confrère de l'aider à construire un diagnostic à partir de résultats d'examens. En général, ils utilisent un premier médium pour visualiser le document (une fenêtre contenant le document) et un deuxième médium pour communiquer à propos du document (le téléphone, une fenêtre pour *chatter*). Ils travaillent sur des parties précises (quelques lignes, un tableau, etc.) et perdent beaucoup de temps à les identifier car ils ne peuvent voir simultanément ce que visualise l'autre interlocuteur. Si les deux professionnels de santé utilisaient un outil d'annotation pour commenter, chacun leur tour, les parties, ils échangeraient ainsi des informations *contextualisées* grâce aux ancres placées dans le document. Ils connaîtraient la partie discutée et le contexte dans laquelle elle se situe.

Les annotations peuvent donc être le support d'interactions sociales (d'échanges liés à l'acte de lecture d'un document et d'écriture des annotations), s'inscrivant dans une activité sociale regroupant plusieurs personnes. Il existe déjà des outils comme *Anchored conversation*[16], *Annotea*[17] qui supportent les conversations synchrones et asynchrones.

[16] Cet outil a été développé au laboratoire FXPAL de Palo Alto [6]

[17] Cet outil a été développé dans le cadre d'un projet du W3C [11]

Les annotations de communication contiennent des traces de l'activité médicale des professionnels de santé.

Le statut d'objet social de l'annotation reste toutefois discutable. Toutes les annotations n'ont pas l'autonomie (*cf.* 3.1.2) et la persistance dans le temps, leur donnant le statut d'objet et d'enjeu de l'interaction. En effet, seules certaines annotations sont conservées définitivement dans le dossier. Par exemple, une fois le patient sorti de l'établissement, les annotations de communication « Penser à changer la perfusion x » n'ont plus d'intérêt et ne seront pas conservées. Á l'inverse, certaines annotations peuvent acquérir le statut d'objet social autonome et permanent, comme une note de suivie qui développe un point précis. Les annotations peuvent donc être le support d'interactions sociales, mais toutes n'ont pas vocation à devenir des objets sociaux.

3.2 Les annotations : des documents particuliers ?

Dans cette dernière partie, nous cherchons à expliquer pourquoi nous considérons les annotations comme des documents particuliers, se rapportant à d'autres documents.

L'annotation et la cible sont perçues par l'annotateur et d'éventuels lecteurs comme deux objets distincts. Lorsque l'annotation est placée dans le document, par exemple sous la forme d'un commentaire dans la marge, le lecteur la différencie du document annoté car la forme (p. ex. l'écriture) et le contenu (p. ex. un point de vue distinct dans le document et l'annotation) peuvent être différents. Il est possible de considérer que cette annotation fait partie du document qui, une fois annoté, devient un nouveau document. Nous pensons plutôt qu'il s'agit de deux documents correspondant à des intentions de communication différentes. L'auteur du document et l'annotateur sont (ou non) deux personnes distinctes. Seul le support d'inscription, dans l'exemple précédent la feuille de papier, est le même. Lorsque l'annotation est extérieure au document, si l'annotateur ajoute un *post-it* sur un document papier, il est plus évident de les considérer comme deux documents car les supports d'inscription sont différents. Sur support numérique, afin de traiter les annotations indépendamment de la cible, nous définissons une annotation comme un document numérique particulier attaché à une cible.

Lors de l'écriture du document [12] et lors de la journée d'étude du 15 janvier 2004 portant sur les documents numériques[18], certains chercheurs se sont interrogés sur le statut de document. Quand un objet devient-il un document ? Ayant défini une annotation comme un document, il nous parait intéressant de nous positionner sur ce thème. [12] affirme que pour acquérir le statut de document, un objet doit remplir deux conditions :

- « l'inscription doit dépasser la communication intime (entre quelques personnes privées) pour devenir légitime »,
- « la légitimité doit s'affranchir de l'éphémère (dépasser le moment de son énonciation) et donc être enregistrée, inscrite ».

En ce qui concerne la deuxième condition, une annotation s'affranchit de l'éphémère puisqu'elle résulte de l'inscription de traces sur un support. L'annotateur interprète le document et selon le point de vue adopté, formule une connaissance qu'il inscrit dans l'annotation. L'information consignée n'est pas éphémère et existe tant que l'annotation existe. Comme pour les documents classiques, le statut de document de l'annotation n'est pas acquis définitivement car elle peut être oubliée, sortie de la conscience collective et être retrouvée et donc relégitimisée [12]. Ceci est d'autant plus vrai pour certaines annotations de communication (*cf*. 3.1.3) qui ne sont pas conservées dans le dossier.

La première condition ne fait pas l'unanimité complète des contributeurs du texte [12]. Certains auteurs pensent que les annotations n'étant pas communiquées, publiées, comme le journal intime ou la liste de course, ne sont pas des documents. « Un journal intime n'est pas un document, sauf si quelqu'un prend l'initiative de le rendre public ou au moins de le communiquer au-delà du cercle restreint des familiers de son auteur. » Pour ces auteurs, un document non communiqué, non publié, ne s'inscrit pas dans une relation sociale car, en s'adressant à lui-même, l'auteur suspend tout rapport avec le monde qui l'entoure. Nous nous positionnons dans la deuxième catégorie d'auteurs qui pensent que les annotations sont des documents car la valeur d'un document « pré-existe à son partage ou son enregistrement ». Pour créer un document, un auteur utilise des connaissances issues de sa compréhension du monde. Suite à l'acquisition de connaissances sur ce monde, il se forge une représentation mentale de ce à quoi doit ressembler le document qu'il rédige. Cette représentation va le guider pour donner une forme au sien. Suite aux relations qu'il entretient avec le monde, il se forge une représentation mentale des connaissances qu'il veut consigner dans son document. Cette deuxième représentation lui permet de décider du

[18] http://h2ptm.univ-paris8.fr/dnumerique/

contenu de celui-ci. L'auteur cherche alors à rendre conforme le document qu'il écrit avec ces deux représentations. On ne peut donc pas dire que la rédaction d'un document peut être coupée de tout rapport social car, même pour les annotations ou le journal intime, c'est grâce à la compréhension du monde qui l'entoure que l'auteur construit les deux représentations utilisées pendant la rédaction. En conclusion, un document n'a pas besoin d'être communiqué à un groupe de personnes non intimes pour acquérir le statut de document. Par contre, tant qu'il n'est pas perçu par un lecteur (p. ex. l'annotateur lui-même), reconnu comme appartenant à un groupe (p. ex. l'annotateur vérifie que son annotation est conforme au groupe des annotations) et interprété (p. ex. l'annotateur interprète ce qu'il a annoté), un objet n'acquiert pas le statut document. Ainsi, les professionnels de santé sont habitués à rédiger et à consulter des annotations dont la généricité de la forme et du contenu leur donne un statut de document.

Nous reformulons donc la première condition de la manière suivante :

- l'inscription doit dépasser le cadre de l'intime plus par le statut du document, instituée, acceptée comme mode de communication usuelle, que par le nombre réel de personnes qui le lisent.

4 CONCLUSIONS ET PERSPECTIVES

Dans la première partie de cette publication, nous avons considéré des documents particuliers, les documents médicaux. Avec les documents papier, les professionnels de santé ont une perception directe du contenu du dossier pour peu qu'ils connaissent les règles d'écriture et de lecture de celui-ci. Avec les documents numériques, ils ont désormais besoin d'un intermédiaire, la machine, pour percevoir ce contenu. Bien que le dispositif de lecture se complexifie, le numérique permet de mettre à disposition du lecteur différentes formes perceptives des documents facilitant sa lecture. En effet, pour peu que le système dispose d'une représentation logique explicite des documents et des données, il pourra fournir aux utilisateurs différentes vues sur le dossier, réagençant les documents et les données de différentes manières.

Dans la deuxième partie, nous avons montré en quoi les annotations font partie du contrat de lecture unissant le(s) auteur(s) aux lecteurs. Nous pensons qu'un outil d'annotation est indispensable en complément d'un système informatisé de gestion des dossiers patient, afin de *contextualiser* les informations non prévues par les concepteurs des documents et qui sont produites lors de la rédaction ou de la lecture du dossier. En effet, les

trois conditions du contrat de lecture ne sont remplies que si les utilisateurs peuvent annoter les documents. Nous avons par ailleurs montré qu'un dossier informatisé contenait essentiellement des formulaires numériques structurés. La dimension de *lisibilité* partagée entre auteur(s) et lecteur(s) ne sera remplie que si le rédacteur et le lecteur annotent, sinon ceux-ci ne disposent pas de moyens suffisants pour laisser des traces du contexte de rédaction. Pour finir, les annotations sont aussi des traces de la *compréhension* que les lecteurs ont des documents et des *relations sociales* entretenues par les acteurs utilisant ces documents. Comme [12], nous proposons une définition sous forme d'équation :

Formulaire annoté = structure et connaissances du concepteur + connaissances du rédacteur + connaissances du lecteur

Cette équation, qui rend compte de l'insertion du lecteur dans le processus constitutif du document, fait de ce lecteur un auteur lui-même. Le formulaire est conçu par le concepteur qui le structure et laisse des traces de ses connaissances dans les intitulés descriptifs. Il est complété par le rédacteur qui laisse des traces de sa compréhension dans les zones à compléter du formulaire et dans les annotations. Le lecteur pourra à son tour laisser des traces des connaissances élaborées à la lecture du document dans les annotations. Comme l'explique [3], la pratique d'annotation est donc une manière, pour le lecteur, de se réapproprier le document, de le réécrire selon l'usage désiré (pour préparer un cours, un document de synthèse, etc.). Il devient ainsi l'« auteur de sa lecture » [3].

De cette synthèse, nous avons dégagé deux points que nous cherchons à intégrer dans le prototype du projet DocPatient : la structure des documents que nous utilisons pour faire des calculs sur les documents et leur contenu et les annotations que nous considérons comme essentielles lors de la rédaction et de la lecture du dossier patient. Nous construisons actuellement la maquette informatique de ce prototype. Elle gère une partie du dossier de spécialité du service de Pédiatrie ainsi que les annotations. Elle va être utilisée dans le service par un groupe de médecins pour améliorer notre modélisation de l'objet annotation ainsi que l'implantation réalisée[19].

Sur le plan théorique, ces premiers travaux ont mis en avant trois besoins : Le premier consiste à définir le sens des différents documents générés à partir de l'ensemble des ressources du dossier informatisé

[19] Dans le cadre de ce projet, nous n'aurons malheureusement pas le temps de nous intéresser aux annotations multimédia, pourtant très pertinentes pour commenter des parties d'images. Nous invitons le lecteur intéressé à se reporter à [16].

(besoin d'une sémantique hypertextuelle [4]). Le deuxième consiste à définir le sens des annotations ajoutées aux documents par le lecteur pour se construire sa propre vision du dossier (besoin d'une sémantique annotationnelle). Le troisième consiste à préciser la notion d'autonomie (*cf.* § 3.1.3) et de persistance dans le temps (*cf.* 3.2) de l'annotation. Nous chercherons à généraliser nos définitions à des contextes plus larges que le dossier patient car les annotations sont utilisées dans de nombreux autres domaines comme la génétique, l'architecture, etc.

RÉFÉRENCES

[1] B. Bachimont. Bibliothèques numériques audiovisuelles. Des enjeux scientifiques et techniques, Document numérique 2 (1999), 3-4, p. 219-242

[2] B. Bachimont. Connaissance et support d'inscription : entre raison graphique et raison computationnelle. 7me Ecole d'été de l'ARCo, Bonas, 10-21 Juillet 2000. http://www.utc.fr/arco/activites/ecoles/Bonas2000/Bachimont.pdf

[3] B. Bachimont. Dossier et lecture hypertextuelle : problématique et discussion. Exemple autour du dossier patient. Les cahiers du numérique, numéro spécial sur l'information médicale numérique, Vol. 2/2, p. 105 à 123 sous la direction de P. Le Beux et D. Boullier, 2001.

[4] V. Brunie. Désagrégation et reconstruction documentaire pour la lecture des hypertextes. Thèse en mathématique et informatique soutenue en 1999. http://vincent.brunie.free.fr/these/index.html

[5] J. Charlet. L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales. Mémoire d'habilitation à diriger des recherches, 2003.

[6] E.F. Churchill, J. Trevor, S. Bly, L. Nelson, D. Cubranic. Anchored Conversations. Chatting in the Context of a Document, Actes de conférence CHI (Human-Computer Interaction), 2000, La Hague, p. 454-461.

[7] L. Denoue. Nouvelles applications pour les annotations électroniques, Rapport technique.

[8] E. Desmontils, C. Jacquin. Annotations sur le Web : notes de lecture, Actes des journées scientifiques Web sémantique, 10-11 octobre 2002, http://www.lalic.paris4.sorbonne.fr/stic/octobre/octobre1/Desmontils.pdf

[9] A. Flory, F. Laforest. Utilisation des documents semi-structurés pour la représentation et le stockage du dossier médical, 2000. http://www.univ-lille2.fr/cerim/tiphad/DocumentsExtranet/lisiRapport2000_07.pdf

[10] K. Khelif, R. Dieng-Kuntz, Annotations pour le domaine des biopuces, Article dans les actes de la conférences Ingénierie des Connaissances 2004, Lyon, 5-7 Mai 2004

[11] Koivunen, R. Swick. "Metadata Based Annotation Infrastructure offers Flexibility and Extensibility for Collaborative Applications and Beyond" http://semannot2001.aifb.uni-karlsruhe.de/papers/1_annotea.pdf

[12] R.T. Pédauque. Document : forme, signe et médium, les re-formulations du numérique. Article. 08 juillet 2003. http://archivesic.ccsd.cnrs.fr/sic_00000511.html

[13] C. Roussey, S. Calabretto, J.M. Pinon. SYDOM, un outil d'annotation pour le Web Sémantique, Journée Web Sémantique 2002, http://www.lalic.paris4.sorbonne.fr/stic/octobre/octobre/apr/Roussey.pdf

[14] F. Therouse. De quelques aspects juridiques sur l'informatisation du dossier patient. Rapport d'avancement de la tâche recherche et action du projet DocPatient RRA/03/TRA3/01, 2003 http://www.laria.u-picardie.fr/EQUIPES/ic/htsc_old/documents/rapport/RRA-03-TRA3-01.pdf

[15] M. Zacklad, M. Lewkowicz, J.F. Boujut, F. Darses, F. Détienne. Forme et gestion des annotations numériques collectives en ingénierie collaborative, Actes de la conférence Ingénierie des Connaissances IC, Laval, 1-4 Juillet 2003, Edition Presses Universitaires de Grenoble, p. 207-224.

[16] E. E. Zsigmond, A. Mille, Y. Prié, J.M. Pinon, Trèfle : un modèle de traces d'utilisation. Article dans les actes de la conférences Ingénierie des Connaissances 2002, Rouen, 28-30 Mai 2002