

Introduction : dans la lignée de Pédaque, pourquoi le document en contexte organisationnel

Maryvonne Holzem, Jacques Labiche

► **To cite this version:**

Maryvonne Holzem, Jacques Labiche. Introduction : dans la lignée de Pédaque, pourquoi le document en contexte organisationnel. Jun 2004, 2004. <sic_00001241>

HAL Id: sic_00001241

https://archivesic.ccsd.cnrs.fr/sic_00001241

Submitted on 8 Dec 2004

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.

Introduction

Maryvonne Holzem¹, Jacques Labiche²

¹ Dyalang FRE CNRS 2787 Université de Rouen² PSI FRE CNRS 2645 INSA & Université de Rouen

Maryvonne.Holzem@univ-rouen.fr

Jacques.Labiche@univ-rouen.fr

1. Intention initiale

1.1. Dans la lignée de Pédaque : pourquoi le document en contexte organisationnel

Dans la lignée de la réflexion pluridisciplinaire sur la notion de document à laquelle nous avons participé avec de très nombreux chercheurs (cf <http://rtp-doc.enssib.fr/pedaque/historique.html>), nous voudrions ici *ébranler* (sic), à notre tour, la notion de document. Nous nous sommes penchés sur un document d'un type particulier, celui qui, produit par et pour un métier, est lié à la vie de l'organisation qui le porte et dont il justifie de l'existence propre.

La notion de *document métier* qui bien que d'apparence plus restreinte que celle de document (dont elle serait l'hyponyme) nous semble tout aussi difficile à cerner que celle de son hyperonyme. Nous ne voulions pas refaire, à une autre échelle, l'important travail de Pédaque, mais plus expérimentalement l'envisager du point de vue de terrains d'investigation spécifiques : autrement dit appréhender le document en contexte organisationnel. De ce point de vue le terme de *document-organisation* semble, sans doute, plus adéquat avec notre démarche. La notion de *métier* renvoyant à l'exercice d'une situation professionnelle plus figée, validée par une formation spécifique, peut paraître insuffisante pour appréhender le document comme ensemble de données intermédiaires consignées sur différents supports, insérées dans un processus de transformation interactionnel et éditorial. Car c'est bien le document dans sa fonction d'intermédiation sur l'axe pragmatique d'une activité commune finalisée que nous convoquons ici. Cette problématique axiologique enchâssée contextuellement devrait nous permettre de mieux cerner les

éléments déclenchant pouvant conduire à des actions. Soulignons cependant que les domaines considérés se situent à la frontière d'une activité fondamentalement monologique, celle de la prise de décision stratégique au sujet du devenir du document [HAB73] et d'une activité dialogique relative à la lignée de réécriture dans laquelle s'inscrit le document [BAK84, RAS01].

L'étude de situations organisationnelles concrètes, pourrait être de ce point de vue riche d'enseignements. Elle devrait nous permettre d'avancer dans la réflexion pluridisciplinaire dans des domaines particulièrement importants quant à leurs enjeux sociaux et économiques : celui des travailleurs sociaux et des ingénieurs brevet.

1.2. Historique de la prise en considération du contexte organisationnel pour le traitement du document

Sans vouloir, nous appesantir sur les difficultés qui ont amené les chercheurs travaillant sur la reconnaissance de formes à contraindre leur approche du document, nous retracerons le chemin qui a conduit à notre rencontre pluridisciplinaire et à la réalisation du travail présenté ici.

La problématique de la rétroconversion du document "papier" à partir de sa numérisation a permis aux chercheurs de mettre au point et tester des outils de traitement d'images et de reconnaissance de formes pour extraire tous les objets présents sur le document puis les archiver dans des bases de données. La réussite scientifique, technique et industrielle de la rétroconversion n'étant pas assurée, il a été nécessaire de contraindre le problème. Certains laboratoires (PSI par exemple) issus des Sciences pour l'Ingénieur *SPI* se sont alors placés dans le contexte du *document métier* conçu dans le cadre d'une organisation avec des objectifs professionnels ; ceci permet en effet d'utiliser des modèles de document et des interfaces homme-machine qui guident la reconnaissance. Afin d'affiner, préciser et opérationnaliser ces modèles et interfaces, il est devenu indispensable de coopérer avec les Sciences Humaines et Sociales *SHS* qui analysent les pratiques professionnelles et les organisations.

Le point de vue de l'organisation (acteurs – méthodes de travail – ressources – normes – corpus...) doit apporter des connaissances sur les contextes de production, de circulation et de consommation des documents liés aux problématiques *métier*. La notion de cycle de vie du document s'ouvre donc aux pratiques des organisations et entreprises.

Modéliser et formaliser les connaissances concernant le document (relié à une collection), le flux (relié à l'organisation) le producteur et l'utilisateur (reliés à une communauté d'intérêt) devient incontournable.

Cette approche intègre plusieurs notions qui nous semblent devoir être spécifiées : la notion d'*intention* et celle de *document situé au sein d'une pratique sociale*.

Nous nous garderons bien de vouloir saturer ici la notion d'intention. Nous nous contenterons de ses acceptions dans le cadre d'une pratique professionnelle. Celles-ci convoquent néanmoins la notion d'*arrière plan* (Searle 1983) comme ensemble des capacités mentales (schèmes, pratiques, compétences, habitudes etc.) qui conditionnent notre perception en ce qu'elles participent à la construction de notre expérience visuelle. Dans le cadre professionnel, cette expérience est, d'une part

fondamentalement partagée (je perçois l'objet, le document, comme quelque chose que je sais devoir percevoir avec d'autres) et elle est d'autre part conditionnée par des faits institutionnels (présupposition d'une norme dont ces faits sont une instance actualisée en contexte).

La notion de document situé dans une pratique sociale inclut, quant à elle, tous les éléments éditoriaux, lexicaux, syntaxiques et sémantiques, constitutifs de cette pratique. Il était alors impératif d'interroger les champs disciplinaires des *SHS* conjointement avec ceux des Sciences "exactes" *SPI*. L'*AS Document et Organisation* s'est donc donnée pour objectif de développer cette collaboration pluridisciplinaire en étudiant des cas réels, en analysant les réussites et les échecs et d'en dresser un bilan.

1.3. Point de vue pluridisciplinaire sur le contexte organisationnel

Nous présenterons succinctement les équipes qui ont initié la recherche puis les apports scientifiques des champs disciplinaires concernés

Pôle SPI :

FRE CNRS 2645 PSI INSA et Université de Rouen : Jacques Labiche, Eric Trupin,
L3i EA 2118 La Rochelle : J-M. Ogier, Remy Mullot
France Télécom R&D : Joël Gardes

Pôle SHS :

FRE CNRS 2787 DYALANG Université de Rouen : Maryvonne Holzem, Sylvie Normand, Valérie Delavigne.
GRESI ENSSIB Lyon : Marie-France Peyrelong, Brigitte Guyot
GRIS Université de Rouen : Martine Blanc, Catherine Peyrard

La sociologie du travail conjointement avec la sociolinguistique analyse les pratiques langagières au travail dans les écrits professionnels ainsi que les pratiques professionnelles implicites et explicites en identifiant si besoin des groupes professionnels grâce à l'analyse de l'organisation de l'entreprise.

En théorisant l'utilisation de corpus, la sociolinguistique, avec le laboratoire Dyalang, s'intéresse à la structuration d'un discours lié à une activité. La notion de corpus renvoie à celle de document ancré dans un contexte, elle est corrélative de celle de genre, c'est-à-dire de types relativement stabilisés d'énoncés élaborés au sein de chaque sphère d'utilisation de la langue [BAK84]. Les sociolinguistes préfèrent alors retenir le terme *sphère d'activité* [GUE93] à celui de *domaine* qui sous-entend l'idée une communauté langagière a priori (celle des chercheurs, des ingénieurs, des techniciens, des gestionnaires etc.) là où les pratiques quotidiennes dans le monde de la recherche comme celui de l'industrie font état d'un partenariat autour de tâches à réaliser. Une sphère d'activité est de fait pluridomainiale [GAM85] elle est un construit en fonction d'un but et non une donnée en fonction d'un découpage préexistant.

L'approche sociolinguistique postule que tout écrit, bien qu'en apparence monologique, est fondamentalement *dialogique* en ce qu'il est attentif au

destinataire, prévient ses critiques, tente de le convaincre, de se justifier, tout en répondant à des raisonnements antérieurs formulés par d'autres locuteurs au sein de la sphère considérée. L'approche descriptive des usages langagiers tente alors de cerner la structure et les marques linguistiques qui constituent les traces de cette situation d'énonciation que Bronckart [BRO87] préfère appeler situation de production et qui réunit l'ensemble des facteurs de création de l'écrit considéré : réalité matérielle, lieu, agent etc. [HOL et WAB01].

Les membres du GRIS (Groupe de Recherche Innovations et Sociétés) se situent dans une perspective qui articule sociologie du travail et sociologie des techniques. Ils s'intéressent donc aux différents salariés dont les pratiques professionnelles sont structurées tout à la fois : par leur place dans la division du travail tant professionnelle qu'organisationnelle et par leur trajectoire professionnelle qui s'insère elle-même dans un parcours de vie. Ils s'intéressent également aux objets techniques dont sont entourés ou non les salariés au cours de leur pratique professionnelle et considèrent que ces objets construisent pour partie leur pratique professionnelle. Le processus de conception et de développement de ces objets ignore souvent leurs utilisateurs. Les professionnels, qui les utiliseront, vont faire avec ou non. Ils vont leur donner sens dans le cadre de leurs pratiques professionnelles. Pour la sociologie du travail le document est une des facettes des écrits professionnels qui sont multiples dans leurs objets, multiples par leurs supports.

La sociologie du travail les analyse dans le cadre des pratiques langagières au travail. Même si l'on s'intéresse à une pratique langagière particulière, il est nécessaire de saisir comment elle s'articule aux autres et pourquoi. Au-delà des particularités propres à chaque individu, pouvons-nous repérer des configurations liées à des groupes professionnels et pouvons-nous en repérer les enjeux pour chacun d'eux? Pouvons-nous repérer les transformations et les enjeux qu'apporte un nouveau type d'écrit, tant dans un groupe professionnel précis qu'entre groupes?

La sociologie des techniques montre, quant à elle, comment les professionnels vont donner sens aux objets techniques, en tant que construit social, dans le cadre de leurs pratiques professionnelles. Elle reprendra les travaux de B. Latour qui mettent l'accent sur le fait que les objets techniques sont des objets sociaux. Qu'ils sont produits dans les relations sociales et produisent en retour du lien social. Toutefois, elle se sépare de Latour en considérant avec G. Simondon que les objets techniques ont également une consistance et une logique propres et rejoint A. Giddens lorsqu'il porte une attention particulière au rôle joué par les transformations des moyens de communication dans la construction de nos sociétés modernes.

Les sciences de l'information autour de l'équipe «Echanges d'information et structuration des organisations» du Gresi Enssib se penchent sur les "écrits au travail" (P. Delcambre), sur le statut de ces documents, sur le passage de "systèmes d'information personnels" aux systèmes d'information collectifs (B. Guyot, M-F. Peyrelong), en utilisant notamment les notions "d'information située" (B. Guyot), et de "convenance". Car les organisations sont aujourd'hui confrontées à la gestion de flux d'information de plus en plus lourds, ce qui implique de formater non seulement les données mais aussi les informations non structurées, et de trouver les

filtres adaptés (en amont et en aval) pour pouvoir travailler ensemble à distance. Elles prennent donc des initiatives pour formaliser et formater les méthodes de travail, et pour centraliser des ressources d'information hétérogènes censées être utiles à chacun. Le problème se pose alors en termes de sélection, de validation, de modalités d'accès à des documents pertinents, en donnant à chacun l'information qui lui convient. Elles doivent alors faire face à une forme de modélisation des besoins et des usages. Encore faut-il pouvoir les identifier et les prendre en compte. Cela pose des questions comme l'implication personnelle dans cette "mise en outils", la répartition des responsabilités en matière de management de l'information.

En adoptant une approche par le document, l'équipe lyonnaise prend celui-ci comme un support d'accord et de représentations partagées, qui inscrit en lui les différentes interventions des acteurs concernés (négociations ou impositions). L'ensemble de la chaîne de production, de mise en forme, de circulation ainsi que les règles et procédures, doit alors être observée pour appréhender le statut du document ainsi que les modalités d'intervention des acteurs (règles éditoriales, règles de circulation, de mise à jour, d'archivage, par exemple). En regardant comment se distribuent les rôles, le suivi et le contrôle.

Dans la communauté des « traiteurs d'images », les algorithmes de reconnaissance de caractères n'ont réellement progressé que lorsque les chercheurs leur ont incorporé des connaissances de plus haut niveau sur la mise en forme matérielle, les vocabulaires utilisés, etc. C'est ainsi que sont nées des équipes « document » en Sciences pour l'Ingénieur. Ces équipes qui se sont ensuite penchées sur la rétroconversion de documents et sur le traitement assisté de formulaires [TRU97] ou de plans [ADA00] ont eu également à intégrer des connaissances de haut niveau sur le contexte de gestion de ces documents [LAB00]. L'échec relatif de cette problématique a débouché sur le besoin de conception de plate formes de traitement informatisé de documents dont le rôle n'était plus l'extraction de tous les objets et informations du document mais uniquement de pouvoir répondre à des requêtes précises sur les contenus textuels et/ou graphiques. Il apparaît en effet qu'on ne saura extraire des informations pertinentes des documents numérisés que pour un lecteur (être social) qui émet une requête "précise" sur le document.

Le système aura alors à sa disposition des outils de traitement d'images, de reconnaissance de formes, des outils linguistiques, des documents et un utilisateur-lecteur muni d'une intention précise qui pourra être déclinée en termes "métier" puis en termes "graphiques".

Les enjeux scientifiques soulevés par la réalisation de ces plate-formes de traitements assistés sont multiples, mais le principal reste la généralité. Comment concevoir une plate-forme, un système qui puisse rendre des services utiles en milieu professionnel pour le traitement de tout type de document métier ? [BAU03]

La complexité de la tâche a nécessité de s'appuyer sur des modèles du document qui permettent de relier un document "tout venant" aux spécificités correspondant à sa catégorie. Ces modèles ont d'abord été constitués de couches superposées (layer) au fond de page (texte, symboles, graphisme,...) puis ensuite composés de modèles graphiques et logiques. Ces derniers permettent une approche "objet" du document [SAI02] et imposent un formalisme qui se traduit par une

structuration logique et une présentation graphique. Le formalisme XML est l'aboutissement actuel de cette approche.

C'est cette modélisation « informatique » de l'objet document qui est ici confrontée aux retours terrains et qui participe à la problématique pluridisciplinaire Document et Organisation.

2. L'approche délicate des sphères organisationnelles : une spécificité du document / organisation.

2.1. Présentation des corpus et des lieux d'étude.

Nous avons décidé d'investir deux sphères organisationnelles distinctes : celle des travailleurs sociaux au sein d'un CIAD (Centre d'Information et d'Accueil Départemental) situé en banlieue rouennaise et celui des ingénieurs brevet dans un cabinet conseil en brevets d'invention (région parisienne)

Dans le premier cas, l'étude porte sur le formulaire de demande d'un revenu minimum d'insertion (RMI). L'enquête sur le terrain a consisté, d'une part, à enregistrer les entretiens entre une Assistante Sociale et un demandeur (entretien durant lequel est instruit le document), puis, d'autre part, à enquêter auprès des mêmes Assistantes Sociales à propos de leur pratique professionnelle relative à cette tâche. Le corpus étudié a donc été fourni par la retranscription de ces deux types d'entretien.

Dans le second cas, le corpus écrit nous a été fourni d'emblée, puisqu'il s'agissait d'un mémoire technique rédigé par une équipe de chercheurs travaillant en reconnaissance de formes et de sa demande brevet correspondante. Celle-ci ayant été rédigée par un ingénieur brevet que nous avons interviewé ainsi que deux de ces collègues, tous rédacteurs confirmés en brevet d'invention. Ces interviews nous ont alors permis d'affiner notre approche du document brevet dans son contexte organisationnel.

Nous n'en dirons pas plus pour l'instant, laissant le soin au lecteur de pénétrer, avec les articles relatifs à ces deux études, dans les cadres organisationnels et les documents pris en compte.

2.2. Une mise en route délicate compte tenu des difficultés d'approche.

Avant de présenter les différents articles de cet ouvrage, nous voudrions faire état des difficultés rencontrées dans nos approches de terrains. Outre les difficultés d'intercompréhension langagière communément rencontrées lorsque cinq équipes de chercheurs travaillant dans des disciplines différentes décident de travailler ensemble, nous avons dû faire face à de réelles difficultés institutionnelles pour, d'une part, approcher ces milieux et, d'autre part nous saisir de *documents* dont la première des particularités est d'être suivis de l'adjectif *internes* autrement dit non diffusés en l'état :

- soit parce que leur vocation est de circuler à l'intérieur de la sphère institutionnelle et qu'il touche à la vie privée des personnes (ce qui est le cas de la demande RMI).
- soit parce qu'il s'agit d'un document en devenir, non destinés à la diffusion commerciale, comme le sont les articles de recherche, et qui doit être protégé des concurrents (ce qui est le cas du mémoire technique). Ou bien qu'il s'agit d'un document non encore protégé par le label de la propriété industrielle qui en fera un brevet et pourra à ce titre le publier.

Côté RMI et CIAD, les démarches : pour obtenir les autorisations administratives de nous rendre sur place, avoir accès à des formulaires déjà remplis et rendus anonymes par les AS, pouvoir obtenir des rendez-vous avec les personnels, nous ont pris plus de trois mois, et demandé beaucoup d'efforts et de persuasion auprès des organismes de tutelles.

Côté Brevet, nous pouvons affirmer sans risque de nous tromper que les choses auraient été tout aussi difficiles voire impossibles, étant donné le côté stratégique des documents en question. Nous avons eu la chance qu'un membre¹ de notre équipe, étant l'un des rédacteurs d'un mémoire technique et demandeur de sa reformulation en brevet d'invention, nous fournisse le corpus idéal et nous permette de pouvoir nous entretenir avec les ingénieurs brevet.

Nous pensons qu'il s'agit sans doute là d'un début de spécificité du document au sein d'une sphère organisationnelle : sa difficulté d'accès.

2.3. Présentation des contributions

Brigitte Guyot, pour la sociologie de l'information, et Sylvie Normand, pour la linguistique, se proposent d'étudier la place du document dans le secteur de la propriété industrielle en s'attachant à la description de la transformation éditoriale d'un mémoire technique en brevet d'invention. Leur approche disciplinaire se croise au point de passage du monde scientifique au monde juridique avec la prise en compte des marques énonciatives et de la stratégie argumentative, d'une part, de la dynamique sociale et d'une analyse sociologique des pratiques des acteurs aux prises avec l'information, d'autre part. Leur étude du document, doublée d'entretiens avec des ingénieurs brevet, montre à quel point celui-ci est indissociable de l'organisation au sein de laquelle il est inséré et contraint pour être transformé en formes opérationnelles juridiquement incontestables.

Martine Blanc-Merigot, Valérie Delavigne, Catherine Peyrard et Marie-France Peyrelong se sont focalisées sur l'instruction de la demande RMI dans un Centre d'Information et d'Action Départemental. Cette approche venue de la sociologie du travail, des sciences de l'information et de la linguistique tentera d'articuler les dispositifs scripturaux (par l'étude du formulaire), pratiques professionnelles (par l'observation in situ des entretiens et l'interview des professionnels) et pratiques discursives (par l'analyse de la retranscription d'entretiens). Cette étude de la mise

¹ L'AS Document et Organisation remercie France Telecom R&D et plus particulièrement Joël Gardes qui ont rendu possible cette étude des brevets.

en mots puis de la mise en document, saisie au niveau du formulaire, met en évidence tout le travail de traduction et de réduction opéré par les assistantes sociales. Il révèle que le document lui-même (le formulaire RMI rempli lors de l'entretien) n'est qu'un élément d'objectivation de la demande et de la personne et qu'il est profondément encapsulé dans les activités de trame (inscrites dans le statut professionnel) et de chaîne (ou d'accompagnement, de traitement de la demande). Il faut donc chercher dans l'interaction, dans l'autour, le contexte, ce qui va donner sens, permettre les décisions, légitimer une situation personnelle en un demandeur puis enfin, en un ayant droit.

Joël Gardes, Dominique Dionisi, Eric Trupin et Jacques Labiche dans le domaine des sciences pour l'ingénieur se sont focalisés sur les concepts et les modèles implicites qui sous-tendent le traitement assisté du document par ordinateur. Les études terrains lors des travaux de l'AS ont conforté l'approche située du document. Les recherches initiales en *SPI* étaient consacrées au document « métier », le travail pluridisciplinaire a interrogé cette notion de métier et a remis en cause les modèles objets et événementiels utilisés jusqu'alors. L'étude présentée considère le document sous son aspect de support d'information destiné à assurer des communications interpersonnelles entre membres d'une même organisation ou d'organisations différentes dans un but partagé. La notion d'infosphère permettant de contextualiser la communication entre utilisateurs, ou communautés d'intérêt, est discutée. L'utilisation de ce concept d'infosphère amène à reconsidérer les modèles classiques des systèmes d'informations et revisite la notion de canal de communication afin de proposer de nouvelles perspectives pour les modèles et les systèmes informatisés de traitement de documents.

Enfin, Christophe Roche a accepté d'analyser avec un regard extérieur nos travaux et plus particulièrement ceux qui concernent le terrain brevet. Son point de vue de spécialiste de l'extraction et de la gestion des connaissances et sa méthodologie d'analyse textuelle permettent ici de faire rebondir la discussion et d'ouvrir de nouvelles perspectives.

3. Bibliographie

- [ADA00] Adam S., Ogier JM., Cariou C., Mullet R., Labiche J., Gardes J. ; "Symbol and character recognition : application to engineering drawings"; International Journal on Document Analysis and Recognition; pp 89-101 (2000)
- [BAK84] Bakhtine, M. ; Esthétique de la création verbale ; Editions Gallimard (1984)
- [BAU03] Baudouin N., Holzem M., Saidali Y., Labiche J. ; "Modélisation des connaissances et construction d'un consensus : apport de la socioterminologie à une plate-forme en traitement d'images"; TIA 2003 : actes des cinquièmes rencontres Terminologie et Intelligence Artificielle, Strasbourg 31 03-1^{er} 04 2003. LIIA-ENSAIS éditeur, pp 54-68 (2003)
- [BRO87] Bronckart J.-P. ; "Interaction, discours et significations"; Langue Française, n° 74 ; pp29-50.(1987)
- [GUE93] Guespin L. ; "Normaliser ou standardiser ?" ; Le langage et l'homme, vol XXVIII, N°4 ; pp.213-222. (1993)
- [HAB73] Habermas J. ; La technique et la science comme idéologie ; Gallimard (1973)
- [HOL01] Holzem M., Wable T. ; "la description terminologique pour rendre compte de l'avancée des connaissances scientifiques en langue française" Recherche en rédaction professionnelle (sherbroke,CA), Vol 1 ; n°1, [[http:// www.usherb.ca/flsh/rtp/default.htm](http://www.usherb.ca/flsh/rtp/default.htm)]
- [LAB00] Labiche J., Ogier JM., Gardes J. ; "Connaissances actionnables et cycle de vie"; Colloque International CIDE'00; Lyon, juillet 2000; pp 217-229 (2000)
- [RASO1] Rastier, F. ; "Arts et sciences du texte" ; PUF (2001)
- [SAI02] Saidali Y., Baudouin N., Trupin E., Holzem M., Labiche J. ; "An interactive application for knowledge acquisition in image processing"; second IASTED International Conference on artificial intelligence and application, Malaga Spain, september 9-12 2002; pp 173-77 (2002)
- [TRU97] Trupin E. : "Du traitement automatique de formulaires"; Revue Document Numérique, Col.Hermès, Vol 1, Nø 1; pp 39-54 (1997)