sciences
et
techniques
de
l'information

INTD

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

INSTITUT NATIONAL DES TECHNIQUES DE LA DOCUMENTATION

MÉMOIRE pour obtenir le

Titre professionnel « Chef de projet en ingénierie documentaire » INTD

niveau I

présenté et soutenu par

Sophie ASSAL

le 30 octobre 2009

Mots-clés d'auteurs et langages documentaires. Réflexions sur la valorisation des revues du pôle éditorial de la Maison René-Ginouvès

Jury
Brigitte Lequeux (CNRS, Maison René-Ginouvès, Documentation)
Fabienne Wateau (CNRS, Maison René-Ginouvès, Direction)
Hélène Zysman (Smile – Open Source Solutions, Consulting, GED/ECM)

Promotion XXXIX

Remerciements

Entreprendre ce type de projet (suivre une formation lourde tout en assumant ses activités professionnelles) représente une sorte de défi personnel. Et, ce mémoire en est, en quelque sorte, la conclusion. Tout personnel qu'il soit, ce projet n'aurait jamais pu être réalisé sans l'intervention et le soutien d'un grand nombre de personnes.

Ma première pensée va à Marion Abélès. Son soutien, ses encouragements et sa confiance m'ont été très précieux. Je tiens à lui témoigner toute ma gratitude.

Mes remerciements et toute ma reconnaissance vont aussi :

- à Pierre Rouillard et Isabelle Déan qui ont fait tout ce qui était en leur pouvoir pour que ce projet de formation se concrétise ;

- à Dominique Michelet qui, bien que peu enthousiaste à l'idée que je fasse une formation aussi longue, a montré de la compréhension et de la solidarité ;

- à Fabienne Wateau, pour son soutien et ses conseils ;

- à Brigitte Lequeux pour sa disponibilité, ses précieux commentaires, la richesse des discussions et son apport de connaissances ; je ne sais si Brigitte s'en souvient, mais elle a fait partie du jury de concours qui m'a permis d'entrer au CNRS. Au moment de l'oral, elle m'a posé plusieurs questions sur les thésaurus. Nous ne savions pas, à ce moment, que nos destins professionnels étaient en train de se sceller…

- à Loraine Marcheix pour ses relectures et ses commentaires ;

- à Hélène Zysman pour ses enseignements, ses commentaires, son écoute et sa patience ;

- à la chaleureuse équipe administrative et à l'équipe pédagogique de l'INTD ;

- à mes collègues (secrétaires de rédaction, documentalistes, archivistes, infographistes…) et collaborateurs (rédacteurs/directeurs de revue, chercheurs, étudiants…) pour leur soutien ; à ceux et celles, en particulier, qui ont accepté de participer à mon enquête et qui ont consacré du temps pour répondre à mes questions ;

- à mes parents, mes amis et mes camarades de promo (Stéphanie, Valérie, Mathieu, Stéphane, Nicole, Nadège, Anthony…) qui m'ont soutenu tout le temps, aidé très souvent et qui ont dû, parfois, jouer de patience avec moi.

Un grand MERCI à tous !

Résumé

Dans ce mémoire, une réflexion sur la valorisation des revues en sciences humaines et sociales a été menée à partir d'une étude de cas au sein du pôle éditorial de la Maison René-Ginouvès (Archéologie et ethnologie), à Nanterre. Cette réflexion a pour point de départ un questionnement sur les mots-clés d'auteurs qu'ils doivent fournir au moment de la publication d'un texte, questionnement qui est devenu le fil directeur de cette étude.

Ce dernier nous a conduit à aborder les langages documentaires, l'indexation des documents, la recherche d'information, sa valorisation (accessibilité, visibilité et diffusion). Une enquête sur la fabrication des mots-clés et les usages qui en sont faits a été réalisée auprès de chercheurs/auteurs. Les résultats de cette enquête ont permis d'avancer quelques propositions en terme de produits et de services documentaires pour valoriser les revues et les informations qu'elles contiennent.

Mots-Clés
MAISON DES SCIENCES DE L'HOMME, MAISON RENE-GINOUVES, NANTERRE, EDITION SCIENTIFIQUE, REVUE SCIENTIFIQUE, SCIENCES HUMAINES ET SOCIALES, LANGAGES DOCUMENTAIRES, RECHERCHE D'INFORMATIONS, VALORISATION DE L'INFORMATION, INDEXATION, WEB, BESOIN D'INFORMATION, PRODUIT DOCUMENTAIRE, SERVICE DOCUMENTAIRE, ENQUÊTE, USAGE

Table des matières

INTRODUCTION 7
Contexte 9
Une Maison des Sciences de l'Homme 9
Le pôle éditorial 9
Revues propres du CNRS 10
Revues soutenues par le CNRS 10
Deux collections de la Maison René-Ginouvès 12
La mise en ligne des revues en SHS 12
Questions de départ et problématique 15
Méthodologie 17
PREMIERE PARTIE : LANGAGES DOCUMENTAIRES. DEFINITIONS, USAGES ET ENJEUX 18
Préambule 20
Quelques notions-clés 20
Quelques repères historiques 22
Typologie des langages documentaires 23
Langages hiérarchisés 23
Langages combinatoires 25
1 L'indexation des documents 28
1.1 Définition 28
1.2 Contexte 28
1.3 Méthode 29
1.4 Indexation automatique 30
2 La recherche d'informations 31
2.1 Définition 31
2.2 Méthodologie 31
2.3 Typologie 32
2.3.1 La recherche factuelle 32
2.3.2 La recherche courante 33
2.3.3 La recherche documentaire 33
2.4 Recherche d'informations et moteur de recherche 33
3 La valorisation de l'information 35
3.1 Quelques définitions 35
3.2 Méthode et moyens 36
4 Focal sur le web sémantique 39
DEUXIEME PARTIE : REVUES EN SCIENCES HUMAINES ET SOCIALES, MOTS-CLES ET RECHERCHE D'INFORMATIONS. ÉTUDE DE CAS 42
5 Une revue en SHS : un système d'informations organisées ? 44
5.1 Quelques repères sur l'organisation des informations dans une revue scientifique 44
5.1.1 Étape de production d'un document dans sa version papier et dans sa version électronique 48
5.1.2 Évaluation du texte reçu 48
5.1.3 Préparation éditoriale de la copie 49
5.1.4 Préparation pour la publication 50
5.1.5 Mise en ligne dans le cas du JSA 50
6 Mots-clés et résumés dans des revues en SHS 53
6.1 Enquête et questionnaire : brève présentation 53
6.2 Mots-clés et résumés : plusieurs points de vue 54
6.2.1 Du côté des auteurs 54

6.2.2 Du côté des directeurs de revues 59
6.2.3 Du côté des secrétaires de rédaction 60
7 Recherche d'informations 62
7.1 Typologie, outils et méthodes 62
7.1.1 Typologie des informations recherchées 62
7.1.2 Faire une recherche d'informations en général 64
7.1.3 Trouver une information dans une revue spécialisée 66
7.2 Index, indexation et tables 68
7.2.1 Le JSA 69
7.2.2 La revue *Paléorient* 74
7.2.3 Des tables de revue 74
TROISIEME PARTIE : VALORISATION DES INFORMATIONS, VALORISATION DES REVUES. QUELQUES PISTES DE REFLEXION 76
8 Indexation et mots-clés : quelques préconisations 78
8.1 Bilan des besoins pour la fabrication des mots-clés 78
8.2 Deux thésaurus à la MAE 79
8.2.1 Les PACTOLS 79
8.2.2 Géoethno 84
8.3 Rapprochement mots-clés et descripteurs 89
9 Recherche d'informations, diffusion et/ou signalement d'informations 91
9.1 Bilan des besoins en termes de recherche d'informations 91
9.2 Des propositions de développement et/ou d'améliorations 93
9.2.1 Bases de données 94
9.2.2 Sites web 97
CONCLUSION 99
BIBLIOGRAPHIE 104
ANNEXES 108
Annexe 1 – Liste des abréviations 109
Annexe 2 – Extrait du Rapport d'activités 2007 de la Maison René-Ginouvès 110
Annexe 3 – Pôle documentaire 112
Annexe 4 – Mots-clés/descripteurs 115
Annexe 5 – Questionnaires 126

Liste des figures

Fig. 1 – La mise en ligne des revues du pôle éditorial de la MAE. 13
Fig. 2 – Copies d'écran des rubriques du JSA. 45
Fig. 3 – Copies d'écran d'une page de présentation d'un texte dans le JSA. 48
Fig. 4 – Schéma acteurs/activités au moment de l'évaluation d'un texte. 48
Fig. 5 – Schéma acteurs/activités au moment de la préparation éditoriale d'un texte. 49
Fig. 6 – Copie d'écran du modèle de document Revues.org pour Lodel. 50
Fig. 7 – Copies d'écran des listes de métadonnées et de styles développées sur le modèle de document Revues.org pour Lodel. 51
Fig. 8 – Scan de la page d'introduction de l'index 1947-1976 du JSA. 70
Fig. 9 – Copie d'écran des index du JSA sur Revues.org. 71
Fig. 10 – Copies d'écran de recherche sur Revues.org pour le JSA. 72
Fig. 11 – Copies d'écran de recherche sur Persée pour le JSA. 73
Fig. 12 – Copies d'écran de recherche sur le site de la revue Paléorient. 74
Fig. 13 – Copies d'écran la page des PACTOLS sur Open Theso web. 81
Fig. 14 – Copie d'écran pour l'accès aux descripteurs des PACTOLS. 81
Fig. 15 – Copies d'écran de la liste des notices indexées par un descripteur des PACTOLS. 82
Fig. 16 – Copie d'écran d'une des notices indexées par le descripteur retenu. 82
Fig. 17 – Copies d'écran d'ADLFI montrant la mention des PACTOLS. 84
Fig. 18 – Copie d'écran de la vue de l'arborescence du thésaurus GeoEthno. 86
Fig. 19 – Copie d'écran de mode interrogation du thésaurus GeoEthno. 86
Fig. 20 – Copie d'écran de la réponse obtenue dans le thésaurus GeoEthno. 87
Fig. 21 – Copie d'écran des informations complémentaires répondant à la requête précédente. 87
Fig. 22 – Copie d'écran de la page « interrogation avancée du catalogue », réseau ethnologie. 88
Fig. 23 – Copie d'écran des résultats obtenus. 88
Fig. 24 – Copie d'écran de la notice complète du résultat 3. 88
Fig. 25 – Copies d'écran des Peuples amérindiens dans les PACTOLS. 90
Fig. 26 – Schéma de valorisation de l'information grâce à l'indexation. 94
Fig. 27 – Schéma de mise en relation mots-clés d'auteurs, descripteurs et métadonnées. 101
Fig. 28 – Schéma du projet Siam (réalisé par Brigitte Lequeux). 103

Introduction

Ce mémoire s'appuie sur une expérience de terrain au sein de la revue (*Journal de la Société des Américanistes* ou JSA [1]) pour laquelle j'assure les fonctions de secrétaire de rédaction depuis janvier 2003 et au sein du pôle éditorial qui a récemment été créé à la Maison Archéologie et Ethnologie (MAE ou Maison René-Ginouvès), à Nanterre (voir annexe 2). Il a pour objectif de proposer des pistes de réflexion sur la valorisation des revues – c'est-à-dire sur la mise en place d'outils et de moyens communs permettant de rendre ces revues et les informations qu'elles contiennent plus visibles et plus accessibles pour le public considéré, soit les chercheurs, les enseignants-chercheurs, les étudiants… – et sur un rapprochement possible entre les méthodes et techniques relevant des sciences de l'information et de la documentation, d'une part, et les méthodes et techniques relevant du domaine de l'édition scientifique, d'autre part, pour développer ces outils, moyens et/ou services.

On ne peut en effet ignorer les nouveaux enjeux qui sont apparus, avec le développement du web en général et la mise en ligne de plus en plus fréquente des revues en particulier, pour sélectionner, classer, trier et rendre accessibles les informations. On ne peut non plus s'épargner une réflexion sur, d'une part, les modalités d'indexation et de recherche d'informations sur les versions électroniques de ces revues, d'autre part, sur l'amélioration de la visibilité des revues et des informations qu'elles contiennent car, sans cela, tout le travail de mise en ligne de ces revues ne servirait à rien. C'est dans ce cadre que s'inscrit ce mémoire.

Pour commencer, je vais rapidement présenter le contexte dans lequel cette réflexion a été menée. Puis j'expliciterai la problématique qui sous-tend ce mémoire. Enfin je décrirai la méthode utilisée pour obtenir des informations.

1. Voir annexe 1 pour toutes les abréviations utilisées dans ce texte.

Contexte

Une Maison des Sciences de l'Homme

La Maison René-Ginouvès, Archéologie et Ethnologie (MAE), est un organisme de recherche et de formation à la recherche qui réunit plusieurs disciplines des sciences de l'homme et de la société : archéologie, préhistoire et protohistoire, ethnologie et sociologie comparative et histoire ancienne et médiévale.

La MAE comprend une unité mixte de service (UMS [2] relevant du CNRS et des universités Paris 1 – Panthéon et Paris Ouest-La Défense), quatre laboratoires de recherche (ou unités mixtes de recherche, UMR), deux formations doctorales. Cette Maison regroupe plus de 300 chercheurs et enseignants-chercheurs permanents (dont un tiers relevant du CNRS) – spécialistes de l'Europe, de l'Asie, des Amériques et de l'Afrique travaillant sur différentes périodes, de la préhistoire à l'époque contemporaine –, 80 ingénieurs, techniciens et administratifs permanents (90% CNRS) et plusieurs centaines de doctorants.

L'interdisciplinarité de ces recherches met en œuvre des équipements et des outils de travail communs. En effet, il existe déjà un pôle documentaire qui associe les deux bibliothèques de la Maison (l'une d'archéologie, l'autre d'ethnologie), le service commun de documentation et le service des archives (voir annexe 3). Un pôle éditorial commence à prendre forme.

Le pôle éditorial

L'USR 3225 de la Maison René-Ginouvès (Archéologie et Ethnologie) accueille, depuis janvier 2009, sept revues (classées ici par ordre d'arrivée dans la Maison) : *Gallia* – composé de *Gallia*, *Gallia préhistoire* et des suppléments, AdlFI ou *Archéologie de la France. Informations* –, *Paléorient*, *Journal de la Société des Américanistes*, *Archives européennes de sociologie*, *Ethnologie Française*, *Revue archéologique* et *Syria*. Depuis 2005, deux collections (Colloques de la Maison René-Ginouvès et Travaux de la Maison René-Ginouvès) ont aussi été créées. Cet ensemble de publications constitue le pôle éditorial de la Maison.

2. Depuis le 1er janvier 2009, cette UMS est devenue l'USR (unité de service et de recherche) 3225.

Revues propres du CNRS

Gallia

Éditeur	CNRS éditions
Équipe	deux directeurs/rédacteurs et cinq secrétaires de rédaction
Site Internet	http://www.revues-gallia.cnrs.fr
Informations complémentaires	Revue entièrement financée par le CNRS et recevant des subventions du ministère de la culture Cette revue d'archéologie nationale, créée en 1942 par le CNRS, a pour mission de publier les recherches et les résultats de fouilles les plus significatifs dans le domaine de l'archéologie en France, de la préhistoire à l'Antiquité tardive. En 1956, est créée *Gallia Préhistoire*. À partir de 1987, création d'une 3e revue, *Gallia Informations, l'archéologie des régions* qui a pour vocation de présenter un panorama complet de l'actualité des fouilles archéologiques en France. D'abord, publiée sur papier de 1987 à 1997, cette dernière revue paraît sur cédérom de 1997 à 2003. À partir de cette date, le choix de la mise en ligne est fait. Parallèlement est défini avec la sous direction de l'archéologie au ministère de la Culture un nouveau produit éditorial, nommé *Archéologie de la France – Informations* (AdlFI), destiné à rassembler et à accélérer la diffusion des informations archéologiques.

Paléorient

Éditeur	CNRS éditions
Périodicité	2 numéros par an
Équipe	1 directeur/rédacteur et 1 secrétaire de rédaction
Site Internet	http://www.mae.u-paris10.fr/Pale
Informations complémentaires	Revue entièrement financée par le CNRS Fondé par Jean Perrot et Bernard Vandermeersch en 1973 avec une aide de la Wenner Gren Foundation, *Paléorient* est devenu en 1975 revue du CNRS. Elle est consacrée à la préhistoire et à la protohistoire en Asie du Sud-Ouest, notamment sur des thèmes communs dans les différents pays du Proche et du Moyen Orient.

Revues soutenues par le CNRS

***Journal de la Société des Américanistes* (JSA)**

Éditeur	Société des Américanistes (association loi 1901)
Périodicité	2 numéros par an
Équipe	1 directeur/rédacteur et 1 secrétaire de rédaction
Edition électronique	Pas de site Internet de la revue http://jsa.revues.org/ http://www.persee.fr/web/revues/home/prescript/revue/jsa
Informations complémentaires	Revue soutenue financièrement et en personnel par le CNRS et financièrement par le CNL (centre national du livre) Fondé en 1895, le *Journal de la Société des Américanistes* est une revue scientifique consacrée aux sociétés et aux cultures amérindiennes envisagées dans la totalité de leur histoire, c'est-à-dire depuis leurs origines jusqu'à nos jours.

***Archives européennes de sociologie* (AES)**

Éditeur	= *European Journal of Sociology*. Éditée par Cambridge University Press
Périodicité	3 numéros par an
Équipe	1 directeur/rédacteur et 1 secrétaire de rédaction à mi-temps
Édition électronique	http://journals.cambridge.org/action/displayJournal?jid=EUR

Ethnologie française

Éditeur	Presses universitaires de France
Périodicité	4 numéros par an
Équipe	1 directeur/rédacteur et 1 secrétaire de rédaction
Edition électronique / Site Internet	http://www.cairn.info/revue-ethnologie-francaise.htm http://www.puf.com/wiki/Ethnologie_française
Informations complémentaires	Cette revue restitue les recherches contemporaines en sciences sociales et participe au débat public sur des questions de société telles que la mémoire des villes nouvelles, l'école, les conflits de mémoires… dans leurs singularités et leurs revendications.

Syria

Éditeur	Institut français du Proche Orient (IFPO)
Périodicité	1 numéro par an
Équipe	1 directeur/rédacteur et 1 secrétaire de rédaction à mi-temps
Édition électronique / Site Internet	http://www.ifporient.org/node/73 http://www.persee.fr/web/revues/home/prescript/revue/syria
Informations complémentaires	Syria. Archéologie, art et histoire, qui paraît depuis 1920, se consacre à l'histoire et à l'archéologie du Proche-Orient sémitique (y compris Chypre) de la préhistoire à la conquête islamique. Elle portait avant 2005 le sous-titre « Revue d'art oriental et d'archéologie ». Elle publie des articles dans toutes les disciplines de ce champ de recherche, archéologie, épigraphie, philologie, histoire, histoire de l'art. La revue publie aussi à l'occasion de courtes notes d'actualité, et consacre dans chaque numéro un épais cahier aux recensions d'ouvrages parus sur le Proche-Orient ancien.

Revue archéologique

Éditeur	Presses universitaires de France
Périodicité	2 numéros par an
Équipe	1 directeur/rédacteur et 1 secrétaire de rédaction à mi-temps
Édition électronique / Site Internet	http://www.cairn.info/revue-archeologique.htm http://www.puf.com/wiki/Revue_arch%C3%A9ologique
Informations complémentaires	La Revue archéologique, créée en 1844, est l'un des plus anciens périodiques scientifiques français. D'abord surtout consacrée au territoire national, elle s'est peu à peu concentrée sur l'Antiquité classique. Seule revue française d'archéologie restée généraliste, elle s'ouvre à toutes les facettes des mondes grec et romain, de l'Ibérie à l'Asie centrale en passant par la Gaule. L'histoire de l'art traditionnelle peut alors côtoyer l'archéothanatologie, la présentation d'un nouveau musée ou d'un site Internet. L'actualité est suivie dans de nombreux comptes rendus de livres, dans le Bulletin annuel de la Société française d'archéologie classique, avec les résumés des conférences mensuelles de cette société, et dans des chroniques bibliographiques.

Deux collections de la Maison René-Ginouvès

Il s'agit des collections « Colloques de la Maison René-Ginouvès » et « Travaux de la Maison René-Ginouvès », mises en place au cours du quadriennal 2005-2008. La première collection a pour objectif la publication des actes des colloques organisés par cette même Maison, chaque année. Ces actes sont publiés dans les onze mois qui suivent la rencontre, soit juste avant la tenue du colloque annuel suivant. La seconde collection édite des monographies ou des travaux collectifs.

La mise en ligne des revues en SHS

Ce pôle éditorial tel qu'il vient d'être décrit est un rassemblement de revues spécialisées dans les domaines de l'archéologie, de l'ethnologie et de la sociologie. En dépit du fait qu'elles sont effectivement des revues en sciences humaines et sociales, elles n'ont pas de statut commun, ni de moyens financiers et humains identiques, ni de thématiques proches. Pourtant, on peut considérer que leurs préoccupations en matière de visibilité et de valorisation se posent en des termes assez semblables : comment attirer plus de lecteurs, voire de nouveaux auteurs ? comment la faire connaître ? comment la diffuser auprès d'un plus grand nombre ?

Ces questions rejoignent certes les politiques de diffusion de chacune des revues, mais pas seulement. Elles soulèvent également d'autres réflexions sur l'accessibilité et la recherche d'informations dans une revue, voire sur la mise en ligne des revues en SHS et sur leur mise en valeur grâce à cette mise en ligne.

Les choix éditoriaux faits par chacune de ces revues en matière d'édition électronique sont variables. Trois revues (*Gallia*, *Paléorient* et *Syria*) sont (ou seront très bientôt) sur Persée ; deux (*Ethnologie française* et *Revue archéologique*) sont sur Cairn, une (*Archives européennes de sociologie*) est en ligne sur le site de Cambridge University Press, enfin une dernière (*Journal de la Société des Américanistes*) est à la fois sur Persée et sur Revues.org. Si les revues en ethnologie et sociologie sont en ligne pour une grande partie, les revues en archéologie commencent seulement à réaliser leur mise en ligne.

L'arrivée sur la toile des revues scientifiques est un événement très important en matière d'accès à l'information. Le portail Persée et la fédération de revues Revues.org sont les deux acteurs français de la mise en ligne des revues dans le domaine des SHS. Ils ont été d'ailleurs choisis, notamment Persée, par plusieurs des revues du pôle. Le premier avait, au moment de sa création (2003), uniquement un but patrimonial, il numérisait les anciens numéros des revues. Depuis quelques temps, les frontières de cet objectif ont été dépassées

pour certaines revues : c'est le cas de *Gallia* et *Paléorient*. Persée met, à présent, en ligne les numéros courants de ces revues.

Revues.org a pour vocation de mettre en ligne les numéros courants. La méthode développée par Revues.org n'est pas de l'ordre de la numérisation avec OCR (comme c'est le cas pour Persée), mais une véritable édition électronique de la revue, grâce à Lodel (logiciel de gestion de contenu ou CMS) qui génère une feuille de style. Si la numérisation est effectivement nécessaire pour les vieux numéros (car peu de revues ont la version électronique de tous les textes parus), elle l'est moins pour les numéros récents pour lesquels il est possible d'utiliser d'autres outils permettant d'enrichir les versions électroniques des textes. C'est pourquoi le JSA a fait le choix d'être sur ces deux plateformes complémentaires.

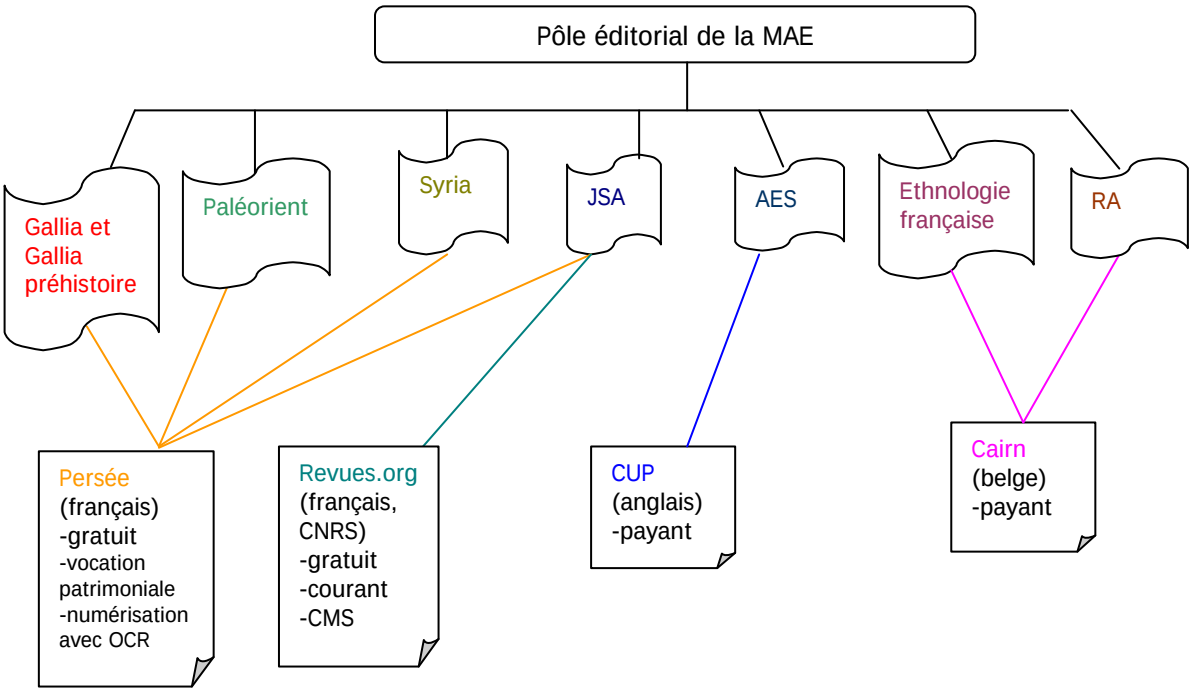


Fig. 1 – La mise en ligne des revues du pôle éditorial de la MAE.

Arrêtons-nous un instant sur le JSA puisqu'il va constituer le socle sur lequel je m'appuierai pour mener une réflexion sur la valorisation des revues. Depuis 2005, les numéros parus de 2001 à nos jours sont sur Revues.org ; ceux parus entre 1895 et 2000 sont sur Persée depuis 2007.

Pour le moment, on peut dire que la mise en ligne du JSA n'est qu'une version électronique de la version papier, voire une version amoindrie de cette dernière puisque, pour les numéros récents, une barrière mobile de cinq ans est en place, empêchant la mise en ligne

intégrale des articles et des notes de recherche pendant cette durée [3]. En revanche tous les autres types de textes (comptes rendus, chronique, nécrologie…) sont en ligne intégralement quelques semaines après leur parution papier. Aucun travail n'a été fait pour enrichir la version électronique. On peut, certes, considérer que, grâce aux outils proposés notamment par Revues.org, les textes mis en ligne sont structurés au niveau de leur forme : l'utilisation des métadonnées Dublin Core et de styles prédéfinis est à l'origine de cette structuration. En revanche, pour ce qui est du contenu proprement dit des textes, on ne peut être satisfait de la façon dont ce contenu est mis en valeur. Pour le JSA, les mots-clés d'auteurs servent à fabriquer des index, mais aucun contrôle sur ces mots n'est véritablement effectué.

D'une façon générale, on constate que la recherche d'informations sur les sites des revues (quand ils existent) ou sur les pages de leur édition électronique (quand elles sont présentes) n'est pas très satisfaisante car pas toujours pertinente, excepté quand on sait très précisément ce que l'on cherche, par exemple l'article de tel auteur sur tel mot du titre. Avec le développement des langages contrôlés (sur lesquels nous reviendrons dans la première partie), le retour d'intérêt pour ces outils depuis plus de deux ans dans le domaine des sciences de l'information et la réflexion menée sur leur évolution en fonction des progrès réalisés pour le web, il devrait être possible de réfléchir à des solutions plus performantes en matière de recherche d'informations dans le cadre de la mise en ligne de ces revues.

3. Cette durée de cinq ans pour la barrière mobile a été définie par le comité de rédaction de la revue pour éviter des pertes financières. En effet, les membres du comité craignaient qu'en mettant en ligne les textes plus tôt, une baisse des ventes au numéro puisse se produire.

Questions de départ et problématique

Revues.org et Persée ont, certes, l'avantage d'offrir aux secrétariats de rédaction des facilités techniques, un soutien humain et, parfois même, financier (Persée notamment), enfin des délais rapides de mise en ligne. En revanche, une fois la mise en ligne effectuée, les rédactions des revues ne peuvent s'épargner une réflexion sur l'accès aux textes publiés et, surtout, aux informations contenues dans ces textes. En effet, aucune des deux plateformes ne s'est véritablement penchée sur l'indexation des revues rassemblées en leur sein, ni sur un plan de classement de ces revues. Elles n'ont pas cette vocation, elles ont pour but d'aider les revues en sciences humaines et sociales à passer en ligne. Pourtant, aujourd'hui, avec un nombre de revues en ligne de plus en plus grand, d'une part, avec le développement des techniques du web, d'autre part, il est important de réfléchir sur la façon d'indexer le contenu des documents mis en ligne.

Cela me paraît d'autant plus important de le faire dans le cadre d'une revue spécialisée que celle-ci peut être une source ou un réservoir de mots servant à l'indexation puisque des spécialistes du domaine sont censés fournir, avec leur texte, des mots-clés. Cela ne signifie pas que ces mots-clés soient tous valables ou même utilisables pour l'indexation, mais ils peuvent servir à suivre l'évolution des concepts d'une discipline et aider à enrichir un vocabulaire déjà mis en place et/ou à développer.

La problématique de ce mémoire va donc se construire autour d'une réflexion sur la valorisation des revues au sein du pôle éditorial de la Maison René-Ginouvès à travers l'indexation des textes et l'amélioration de la recherche d'informations dans les versions électroniques des revues. En effet, ma réflexion a été menée, d'une part, à partir des outils que la documentation propose pour permettre une meilleure valorisation des revues, d'autre part, à partir des revues elles-mêmes et de la façon dont elles pouvaient contribuer au développement de ces outils documentaires, notamment les langages contrôlés.

J'essaierai de répondre à plusieurs questions :

Les rédactions doivent-elles indexer les articles publiés dans leur revue ? si oui, doivent-elles utiliser des langages contrôlés ? Comment les utiliser ? Lesquels ?

Qu'en est-il des mots-clés demandés aux auteurs ? à quoi servent-ils selon les secrétaires de rédaction et selon les auteurs ? Doit-on les utiliser pour envisager une indexation ? Peuvent-ils venir enrichir un langage déjà existant ? Doit-on orienter la fabrication de mots-clés en indiquant aux auteurs ce qu'il serait souhaitable d'y trouver ?

Avant le passage en ligne des revues, celles-ci développaient des outils papiers facilitant les recherches d'informations sur l'ensemble des numéros parus, comme les tables et les index. Aujourd'hui, ces outils n'ont plus besoin d'exister sous leur forme papier, mais qu'en est-il d'une version électronique ?

Pour tenter de répondre à ces questions, j'ai donc choisi de prendre comme fil directeur, tout au long de ce travail, les mots-clés donnés par les auteurs.

Méthodologie

La méthode utilisée pour aborder ces questions est celle de l'observation participante, pour reprendre une méthode chère aux anthropologues, puisque je suis moi-même secrétaire de rédaction au sein de ce pôle, donc à la fois en position d'actrice, car menant les mêmes activités que les autres secrétaires de rédaction et ayant des pratiques équivalentes aux leurs, et d'observatrice, car confrontée aux mêmes problèmes et, donc, à l'écoute de toutes les suggestions possibles pour tenter de proposer des solutions à ces problèmes.

Pour mieux cerner ces questions, j'ai donc mené, d'une part, une petite enquête auprès d'auteurs (chercheurs, enseignants-chercheurs, étudiants…) au sujet des mots-clés et auprès de secrétaires de rédaction et directeurs/rédacteurs de revues sur les mots-clés également, mais aussi sur la valorisation de leur revue. D'autre part, j'ai tenté de rassembler un certain nombre d'observations que j'ai pu faire tout au long de mes années de travail au sein du JSA et de la MAE. Je me suis également servie des discussions que j'ai pu avoir, de façon formelle ou informelle, avec des collègues, des responsables de services, des prestataires (comme les personnes travaillant pour Revues.org, Persée…). Ces observations et/ou discussions m'ont apporté des informations aussi bien sur la mise en ligne du JSA et des autres revues que sur la collaboration au sein de la MAE entre le pôle documentaire et les revues.

Pour l'enquête, j'ai mis sur pied deux questionnaires (voir annexe 4) : l'un destiné aux auteurs de textes publiés dans une revue ; l'autre pour les professionnels de l'édition travaillant dans ces revues (secrétaires de rédaction, directeurs et/ou rédacteur de revue). Le premier portait sur les méthodes adoptées pour la recherche d'informations et sur la façon d'élaborer les mots-clés ; le second abordait l'utilisation et le traitement des mots-clés dans les revues et la valorisation de ces publications.

Dans certains cas, ces questionnaires ont fait l'objet d'entretiens. Dans tous les cas, ils ont été dépouillés en suivant trois grands axes : les mots-clés, la recherche d'informations et la valorisation des revues. Les résultats sont présentés dans la partie 2 de ce mémoire.

Ce mémoire est construit en trois parties : la première est théorique et permet de revoir quelques notions-clés concernant les langages documentaires et les mots-clés ; la deuxième dresse le tableau des résultats de l'enquête présentée ci-dessus ; enfin la troisième propose une réflexion autour de la valorisation des revues, tenant compte de ce qui existe et de ce qui pourrait se faire.

Première partie :
Langages documentaires.
Définitions, usages et enjeux

Pour commencer, rappelons brièvement le cadre théorique dans lequel se déploient ces notions de « mots-clés », « descripteurs », « métadonnées »… Je propose d'y entrer par le biais des langages documentaires [4] car ils occupent une place fondamentale dans le traitement documentaire, la recherche, l'accessibilité et la diffusion de l'information.

Plusieurs façons de définir la notion de « langage documentaire » sont proposées. Nous en retiendrons deux. La première est tirée de l'ouvrage de Jacques Maniez (2002) :

> Codes sémantiques de représentation des sujets permettant à un système documentaire de repérer les documents par une formulation rigoureuse de leur contenu, et aux utilisateurs d'ajuster leurs interrogations à ces formulations. ([12] Maniez, p. 207)

La deuxième provient du *Vocabulaire de la documentation* (2004) :

> Langage contrôlé et normalisé utilisé dans un système documentaire pour l'indexation et la recherche. Un langage documentaire permet de représenter de manière univoque les notions identifiées dans les documents et dans les demandes des utilisateurs, en prescrivant une liste de termes ou d'indices, et leurs règles d'utilisation. ([1] Boulogne)

Retenons, de ces définitions, plusieurs termes : représentation des sujets ou des notions dans les documents, utilisateurs de ces documents, interrogations ou demandes en matière d'information, contrôle, règles et normalisation.

Il apparaît qu'un langage documentaire intervient tant au niveau de l'indexation pour représenter le contenu d'un document qu'au niveau de la recherche d'informations pour répondre aux besoins des utilisateurs. Les langages documentaires sont donc des outils permettant de rendre compte des informations contenues dans un document et/ou une question d'un utilisateur. Ils sont utiles et peuvent être utilisés à chacune des trois grandes étapes de la chaîne documentaire (la collecte de l'information, des documents ; le traitement matériel et intellectuel ; la diffusion et la recherche de l'information). Je vais ici considérer les deux dernières étapes car la première met, en quelque sorte, en jeu les mêmes pratiques de recherche que celles définies dans la troisième. En revanche, je propose de distinguer la recherche et la diffusion de l'information car, bien que dépendante l'une de l'autre, elles correspondent à deux points de vue dans la façon d'accéder à l'information. Pour mieux comprendre l'utilité (à quoi cela sert ?) et l'usage (comment on s'en sert ?) des langages documentaires, je reprendrai chacune des grandes étapes de la chaîne documentaire.

---

4. Nous emploierons, dans ce mémoire, de façon indistincte les expressions « langages documentaires » et « langages contrôlés ». Pour plus d'informations, se reporter à Maniez ([8] Maniez).

Préambule

Quelques notions-clés

Après avoir défini ce qu'on entendait par langage documentaire et/ou contrôlé, donnons quelques définitions [5] de certaines notions dominant ce sujet.

Un **document** est un « support d'information, quel qu'il soit, des données enregistrées sur ce support et de leur signification, servant à la consultation, l'étude, la preuve ou la trace, etc. : livre, échantillon de parfum, tissus, film, etc. Le tout constitue une unité autonome » ([1] Boulogne). Pour le dire autrement, un document correspond à la fois à une forme matérielle (un support) et à un fond (le contenu, les données, les informations).

On peut donc dire que n'importe quel objet peut acquérir le statut de document. Si c'est le cas, la difficulté est de savoir comment le retrouver et comment accéder aux informations qu'il contient. Pour cela, il faut pouvoir identifier une ou plusieurs caractéristiques de cet objet pour le distinguer des autres. On établit alors des critères de recherche comme la taille, la marque, la couleur… Il s'agit de critères objectifs de description sur la forme. Pour un document textuel par exemple, cela correspond à la description bibliographique avec les champs auteur, titre, date, éditeur… Il ne s'agit là que de la description formelle d'un document et non de son contenu. Or, la plupart du temps, lorsque l'on recherche un document, on cherche à l'identifier à partir de son contenu, et cette opération de recherche est rendue possible grâce à l'indexation des documents.

Avec la création des bases de données documentaires dans les années 1960, il a été nécessaire de distinguer les **documents primaires** (document original ou document source) des **documents secondaires** (qui permettent de signaler le document primaire sous forme de fiche de description…).

En raison de la généralisation de l'informatisation, il est nécessaire de préciser ce qu'on entend par **document électronique**. Il s'agit d'un « document (fichier de données et/ou de programme) sur support électronique nécessitant un matériel pour être consulté. Il bénéficie des possibilités de l'informatique : reproductibilité (sur plusieurs écrans en même temps), modification selon le type de code, transmissibilité via les réseaux de

5. Ces définitions sont en grande partie tirées du *Vocabulaire de la Documentation* ([1] Boulogne). Certaines d'entre elles ne tiennent pas toujours compte des évolutions techniques de ces dernières années. Elles ont néanmoins le mérite de donner les informations de base sur ces notions-clés.

communication. On distingue trois modes principaux de représentation de ces documents électroniques : mode caractère, mode image, mode vectoriel » ([1] Boulogne).

Qu'il soit électronique ou non, un document contient des informations. Selon le *Journal officiel* (du 17 janvier 1982), une **information** est un « élément de connaissance susceptible d'être représenté à l'aide de conventions pour être conservé, traité ou communiqué ». Dans le *Vocabulaire de la Documentation* ([1] Boulogne), est ajouté à cette définition : « toute information se caractérise par un contenu (signifiant), un signifié et une forme. L'information est déterminée : pour une représentation analogique, par un vecteur physique de transport (signal électrique, ondes acoustiques et optiques, flux magnétique, radio, etc.) et des paramètres associés à ce vecteur (amplitude, fréquence, format, etc.) ; pour une présentation numérique, par ensemble de bits ».

De la même façon que, pour le document, il a été nécessaire d'introduire la notion de « document électronique » qui va de pair avec le déploiement de l'informatisation, il nous faut ici évoquer la notion d'**information structurée**. Il s'agit en effet d'une « information utilisable directement par un ordinateur pour effectuer un calcul. Ces calculs peuvent être variés : opérations arithmétiques (additions, multiplications…), comparaison (évaluation d'une requête booléenne par rapport à un document par exemple) […]. L'objet du traitement de l'information non structurée est de rendre des données (texte, images, sons…) calculables. Ainsi ce n'est pas la nature de l'information, mais l'utilisabilité d'une donnée qui marque la frontière entre structuré et non structuré » ([7] Livre Blanc. *Valorisation de l'information non structurée*). On peut considérer que 99% de l'information est non structurée ([9] Rais, p. 3).

L'information structurée ou non peut être déclinée en plusieurs formes. Elle peut être présentée sous la forme de données, de thèmes, de termes, de mots-clés…

Une **donnée** est la « représentation d'une information sous une forme conventionnelle destinée à en faciliter son traitement (JO du 17/01/1982, arrêté du 22/12/1981 relatif à l'enrichissement du vocabulaire de l'informatique) » ([1] Boulogne).

Un **thème** renvoie à un « sujet, matière, spécialité, domaine de la connaissance permettant le regroupement d'un ensemble de notions par centres d'intérêt » ([1] Boulogne).

Un **terme** correspond à un mot, une expression ou un symbole « employé pour représenter une notion à l'intérieur d'un domaine. Les termes se caractérisent par leur composition (terme précoordonné ou composé, uniterme), leur relation avec d'autres termes en fonction de leur proximité sémantique (terme associé, terme générique et terme spécifique, terme partitif, terme équivalent). Deux termes de sens distincts sont dits homonymes s'ils ont la

même forme graphique (homographe) ou la même forme phonique (homophone). Deux termes distincts de même sens sont dits synonymes ou termes équivalents » ([1] Boulogne).

Un **mot-clé** est un mot ou une expression « choisi généralement dans le titre ou le texte d'un document pour en caractériser le contenu et en permettre la recherche. Il constitue un point d'accès. Il est à distinguer d'un descripteur qui est un terme normalisé dans un thésaurus » ([1] Boulogne).

Tous ces types d'information ont également une représentation électronique que l'on appelle **métadonnée**, c'est-à-dire un « ensemble structuré de données créées pour fournir des informations sur des ressources électroniques. Elles peuvent remplir différentes fonctions : a) gestion des ressources décrites (suivi du cycle de vie : création, modification, archivage) ; b) informations sur le contenu de la ressource pour en faciliter la découverte, la localisation, l'accès ; c) suivi de l'utilisation et du respect des droits et conditions d'utilisation associés à la ressource. Elles peuvent être exprimées dans le même format technique de codage que celui de la ressource qu'elles accompagnent et être disponibles en même temps qu'elle. Les métadonnées peuvent être écrites actuellement selon plusieurs standards : RDF *(ressource description framework)*, TEI *(text encoding initiative)*, syntaxe « méta » HTML et Dublin core, DTD EAD *(encoding archival description)* » ([1] Boulogne).

Quelques repères historiques

Après ces quelques définitions, il est bon de poser des repères historiques [6] pour mieux situer l'évolution de ces pratiques et de la réflexion qui les sous-tend.

Au XIX^{e} siècle, le monde des bibliothèques, et de la documentation, est marqué par les grandes classifications (CDU, Dewey…). On organise, on range, on classe les connaissances pour mieux les retrouver. Au début du XX^{e} siècle, avec l'accroissement des publications sous forme d'articles, il devient nécessaire de développer des documents secondaires comme les listes de vedette matière.

À la fin des années 1950, le développement de l'informatique favorise la création de thésaurus. Mais c'est surtout dans les années 1970 que les langages documentaires vont connaître un véritable développement qui se fera conjointement avec l'informatique. Ces techniques informatiques vont, en effet, permettre d'effectuer des opérations et/ou calculs complexes dont le but sera de mieux contrôler les langages et d'aider à la recherche d'informations en proposant notamment le croisement de plusieurs critères. Cette possibilité

6. Pour rédiger cette partie, je me suis beaucoup inspirée de l'article de Bruno Menon ([13] Menon).

de combiner des mots pour faire une recherche va de pair avec l'élaboration de langages dits combinatoires post-coordonnés dans lesquels chaque terme et/ou mots-clés est autonome, ce qui permet d'éliminer en partie le problème de l'homogénéisation syntaxique.

Dans les années 1980-1990, l'indexation automatique sur le texte intégral et la recherche également en texte intégral connaissent un développement qui n'a cessé de croître depuis lors. Ces deux techniques viennent concurrencer les langages contrôlés, même si on ne peut ignorer que la recherche en texte intégral est souvent accompagnée d'une recherche par mots-clés.

On pourrait croire qu'avec le développement d'outils de traitement automatique du langage naturel (TALN), l'indexation et les langages documentaires vont être de moins en moins utiles et utilisés. Pourtant, il semble que, dans de nombreux domaines scientifiques, on ne soit pas tout à fait prêt à abandonner ces techniques. Le cas du portail TermSciences, produit par l'Institut National de l'Information Scientifique et Technique (INIST), en est une illustration. Il s'agit en effet d'une base de données terminologique résultant de la fusion de plusieurs ressources terminologiques produites par différents organismes de recherche ([6] Khayari). De plus, avec le développement du web sémantique, on peut se demander si l'indexation et les langages documentaires ne vont pas connaître un nouveau regain d'intérêt.

Typologie des langages documentaires

Les langages documentaires répondent à plusieurs fonctions ([12] Maniez) :

- Une fonction normalisatrice qui permet une harmonisation entre les termes de recherche et ceux de l'indexation, du contenu ;
- Une fonction d'ajustement qui permet à l'usager de s'approcher au plus près de son thème de recherche par appariements successifs, grâce aux liens sémantiques et/ou linguistiques entre les termes d'un langage contrôlé.
- Une fonction d'aide à la sélection des documents grâce à l'indexation proposée dans les notices.

Langages hiérarchisés

Les classifications

Il s'agit d'un « langage documentaire permettant l'organisation d'un ou plusieurs domaines de la connaissance en un système ordonné de classes et de sous-classes. Celles-ci ainsi que

leurs relations peuvent être représentées par les indices d'une notation. Ces indices sont explicités par un libellé textuel. Différents critères peuvent être choisis pour différencier les classifications : le mode de division et des structurations (classification à facettes) ; la notation décimale ou alpha-numérique ; la couverture des domaines de la connaissance traités (classifications encyclopédiques telles la CDU et la classification Dewey, spécialisée ou sectorielle) ; leur structure et organisation (classification hiérarchique), etc. » ([1] Boulogne).

C'est au XIX^e^ siècle que les grandes classifications apparaissent aussi bien en Europe qu'aux États-Unis : la classification décimale (1876) de Melvill Dewey, la classification décimale universelle (1895) de Paul Otlet, l'expansive classification de Amni Cutter pour la bibliothèque du Congrès…

D'après Bruno Menon ([13] Menon, p. 18), « elles résultent du foisonnement éditorial de la fin du XIX^e^ siècle, qui provoque un accroissement considérable des collections. Conçues pour permettre un arrangement rationnel des ouvrages sur les rayons, elles facilitent l'accès thématique aux ouvrages, mais simplifient également la gestion de ces accroissements. De plus, leur système de notation (en particulier la notation décimale) leur confère des capacités d'aménagement et d'extension qui les rendent aptes à suivre les évolutions amenées par les progrès scientifiques ».

Elles sont utilisées dans de nombreuses bibliothèques, même si elles présentent une certaine rigidité. Elles n'offrent pas une richesse pour l'indexation en raison du fait que les relations sémantiques prennent uniquement la forme d'une inclusion. De plus, elles sont très bien pour classer des ouvrages, monographies, mais deviennent inutiles pour toutes unités documentaires fractionnées (comme un article de périodique). Elles aident en effet à classer un ouvrage par son thème principal, mais ne permettent pas de proposer des thèmes secondaires ou d'autres points de vue ou encore des idées émergeantes.

Les plans de classement

Il s'agit d'un « document qui présente une classification de manière ordonnée, en faisant apparaître la signification donnée à chaque indice et, le cas échéant, les relations entre les classes. Le plan de classement peut également inclure des recommandations ou des consignes quant à l'utilisation de la classification, par exemple, pour l'ordre de rangement des documents sur les rayonnages » ([1] Boulogne). Ces outils présentent les mêmes limites que les classifications dont ils rendent compte.

Langages combinatoires

Listes d'autorités

Ce sont des listes « de termes normalisés, soit des mots matières, soit des noms propres, qui doivent être obligatoirement et nécessairement utilisés dans l'indexation » ([1] Boulogne).

Lexique

C'est un « ensemble de termes d'un domaine de la connaissance, classés dans l'ordre alphabétique » ([1] Boulogne). Ces termes peuvent être appelés mots-clés. Il s'agit de langages documentaires à structure combinatoire car c'est par la combinaison de plusieurs mots-clés que l'on peut rendre compte du contenu d'un document.

Liste vedette ou liste de points d'accès

Une vedette est un point d'accès, c'est-à-dire un « terme choisi pour représenter, dans un index, un élément ou une notion contenue dans un document. On distingue : le point d'accès principal représentant l'essentiel du sujet que le point d'accès exprime (vedette principale) ; le point d'accès subordonné (sous-vedette) associé à un autre point d'accès pour indiquer qu'il lui est subordonné ou qu'il le modifie ; le point d'accès matière représentant synthétiquement le sujet ou l'un des sujets d'un document (vedette matière) ; le point d'accès auteur renvoyant à l'auteur du document etc. Le terme de vedette est plus fréquemment utilisé pour les catalogues de bibliothèque » ([1] Boulogne).

Il existe plusieurs listes vedette : la liste de la Bibliothèque du Congrès américain (LCSH), le Répertoire de vedette matière de l'université Laval (RVM) ou encore Rameau. Ce sont des outils très utilisés par les catalogueurs ([13] Menon, p. 19).

Thésaurus

Il s'agit d'« une liste organisée de termes normalisés (descripteurs et non-descripteurs) servant à l'indexation des documents et des questions dans un système documentaire. Les descripteurs sont reliés par des relations sémantiques (génériques, associatives et d'équivalence) exprimées par des signes conventionnels. Les synonymes (non-descripteurs) sont reliés aux descripteurs par la seule relation d'équivalence. On peut distinguer les thésaurus en fonction du mode de regroupement des termes (thésaurus à facettes) [...] ; des domaines de connaissances couverts (thésaurus spécialisé ou sectoriel, thésaurus encyclopédique) » ([1] Boulogne).

D'après Menon ([13] Menon, p. 19), on peut considérer qu'un thésaurus a trois caractéristiques qui permettent de le définir : c'est un langage structuré, contrôlé et

combinatoire. De plus, « la logique de recherche booléenne qui met à profit cette dernière caractéristique est remarquablement adaptée aux opérations de base effectuées par l'ordinateur, d'où un succès immédiat et durable de ces langages documentaires en contexte informatisé » ([1] Boulogne).

Terminologie

C'est un « ensemble de termes propres à un domaine dont la signification fait l'objet d'un consensus chez les spécialistes de ce domaine » ([1] Boulogne).

Ontologies

Les ontologies fonctionnent sur un principe relationnel, des « associations », c'est-à-dire de liens sémantiques entre le sujet de recherche, soit le mot-clé entré dans un moteur de recherche sémantique, et les résultats de la recherche qui sont liés, non plus à la forme syntaxique du mot-clé, mais à sa signification. Cette notion est liée à une évolution à venir du web, que l'on qualifie de web sémantique. ([3] Desfriches Doria)

La littérature sur ce sujet est abondante [7]. La définition la plus répandue est celle de Tom Gruber qui désigne une ontologie comme une spécification explicite d'une conceptualisation. ([3] Desfriches Doria). Pour Véronique Mesguich ([21] Mesguich), « il s'agit de systèmes d'organisation des connaissances, applicables à un ensemble d'objets et de concepts, pourvus de relations et de règles d'inférences entre les concepts ». Bruno Menon ([20] Menon) explique que « les ontologies doivent servir de cadre de contrôle et de référence pour l'expression des métadonnées et leur interprétation par des applications "intelligentes". Elles jouent donc dans ce cadre un rôle similaire à celui des langages documentaires dans les systèmes d'informations bibliographiques ». Pour Charlet, Bachimont et Troncy ([17] Charlet), il s'agit d'un « ensemble structuré de relations, principalement l'hyperonymie des objets du domaine ».

Il semblerait que « de prime abord, la frontière entre **thésaurus** – ensemble de termes normalisés fondé sur une structuration hiérarchisée et où les termes sont organisés de manière conceptuelle et reliés entre eux par des relations sémantiques – et **ontologie** – conceptualisation organisée en vue d'aboutir à un objet formel, où les relations entre les concepts sont sémantiques mais formalisées – semble très mince. En effet, sans entrer dans les détails nous pouvons dire dans un premier temps que tous les deux sont des ressources terminologiques construites dans différents domaines de la connaissance, basées sur des relations à forte dominance hiérarchique et conçues comme un réseau sémantique.

7. Se reporter au mémoire de Marcheix ([18] Marcheix).

Cependant, il existe des spécificités propres à chacun qui nous amène à penser qu'un thésaurus peut servir de base à une ontologie. Les ontologies informatiques, phénomène relativement récent, sont considérées comme les successeurs logiques des thésaurus, comme l'étape supérieure tant d'un point de vue linguistique qu'informatique et les possibilités que paraît ouvrir la création d'une ontologie sont nombreuses et tentantes : l'amélioration de la recherche documentaire, l'automatisation de l'indexation ou encore l'accessibilité et la visibilité du produit sur le web » ([18] Marcheix).

1 L'indexation des documents

L'indexation des documents grâce à des langages documentaires ou des mots-clés n'est pas un processus nouveau [8]. Aujourd'hui, avec la multiplication des documents électroniques et le développement des moyens informatiques pour accéder rapidement à l'information, elle occupe une place centrale car, idéalement, en décrivant le contenu d'un document, elle permet à tout un chacun de le retrouver, ce qui n'est pas négligeable !

1.1 Définition

D'après la définition donnée dans le *Vocabulaire de la documentation*, l'indexation, au sens documentaire, est le processus destiné à représenter, au moyen des termes ou indices d'un langage documentaire ou au moyen des éléments d'un langage libre, les notions caractéristiques du contenu d'un document (ressource, collection…) ou d'une question, en vue d'en faciliter la recherche, après les avoir identifiées par l'analyse ([1] Boulogne).

L'indexation peut être libre – c'est-à-dire que l'indexation se fait […] grâce à des notions et concepts choisis librement par la personne qui indexe – ou contrôlée – c'est-à-dire que l'indexation se fait en langage contrôlé et normalisé par sélection des termes d'indexation dans un langage documentaire de référence (lexique, thésaurus) ([9] Raïs).

La problématique fondamentale de l'indexation peut se résumer ainsi : comment représenter le contenu d'un document ? comment le repérer et le retrouver ? Représenter le contenu d'un document revient à dégager des thèmes, des sujets, des aspects, des données… à déterminer les informations contenues dans ce document. La finalité d'une telle opération est de faciliter l'accès au contenu d'un document ou d'un ensemble de documents à partir d'un thème, d'un sujet, d'un concept ou d'une combinaison de thèmes, sujets ou concepts.

1.2 Contexte

Vers 1995, le document électronique est devenu incontournable. De véritables fonds documentaires numériques ont pu être constitués. L'information n'y est cependant pas toujours structurée, au sens où elle n'est pas toujours utilisable.

D'après Raïs [9], avec la création de base de données documentaires dans les années 1960 et la nécessité de distinguer les documents primaires des documents secondaires, on est passé d'un objectif de conservation/identification à un objectif d'identification/recherche

8. Se reporter à l'ouvrage de Timimi et Kovacs ([11] Timimi).

/gestion, ce qui a conduit à développer des normes de structuration, d'identification et de description des documents. La référence bibliographique n'étant plus suffisante, des bases de données en texte intégral ont été développées, le texte étant accompagné de notices documentaires composées de métadonnées. En effet, on comprend aisément que plus l'information est structurée et organisée, plus l'accès à cette information devient facile. Le fait de proposer du texte intégral n'épargne en rien la création d'une notice descriptive complémentaire, comprenant à la fois des informations bibliographiques et d'analyse du contenu, qui est une aide à la consultation même du texte.

1.3 Méthode

Dans les étapes de la chaîne documentaire, l'indexation apparaît dans la phase du traitement du document, appelée aussi analyse documentaire ou description du contenu ([10] Serre). Remarquons que, dans la définition du mot « document » telle que nous l'avons donnée plus haut – ensemble d'un support d'information, quel qu'il soit, des données enregistrées sur ce support et de leur signification, servant à la consultation, l'étude, la preuve ou la trace, etc. : livre, échantillon de parfum, tissus, film, etc. Le tout constitue une unité autonome. ([1] Boulogne) –, il apparaît une distinction entre support et contenu (mentionnée grâce à la présence de termes tels qu'information, données, signification).

Pour réaliser cette phase, le support matériel et le contenu intellectuel du document doivent être séparés car les opérations qui permettent de décrire ce contenu sont différentes, mais tout aussi importantes, de celles qui donnent des informations sur le support. Ces dernières relèvent de la description bibliographique ou catalogage : il s'agit de l'ensemble des données bibliographiques décrivant un document, servant à son identification et pouvant être structurées en suivant les prescriptions des normes nationales et internationales. La description bibliographique est utilisée dans l'élaboration des notices bibliographiques et catalographiques ([1] Boulogne).

D'après Alexandre Serre, dans son *Introduction à l'indexation*, la description du contenu découle de l'analyse documentaire d'un document. C'est en effet cette analyse qui permettra de construire un résumé et/ou une indexation. Réalisée non pas grâce à une lecture *in extenso* du document, mais grâce à toutes les informations relevées dans des passages significatifs, comme le titre, la table des matières, l'introduction…, ce type d'analyse consiste à : 1/ repérer, sélectionner les concepts, thèmes ou données caractéristiques du contenu d'un document ; 2/ les extraire et les exprimer sous forme de mots-clés. Le problème de ce type d'analyse est qu'elle peut varier selon les personnes qui la font, selon l'époque où elle est réalisée, selon le document et les techniques (notamment le langage) utilisées.

Néanmoins, à partir de cette analyse, on peut donc produire des résumés, des extractions de données (chiffres, citation, extraits significatifs…) et une indexation.

Un résumé est une présentation, plus ou moins concise, du contenu d'un document et peut prendre plusieurs formes : il peut être critique ou analytique. En documentation, il est question de résumés analytiques qui sont soit indicatifs (très court, à peu près 5 lignes), soit informatifs (plus long, entre 15 et 20 lignes).

L'analyse documentaire doit donc également permettre de dégager l'information qui y est contenue et de traduire cette information à l'aide de mots, tirés du langage naturel, les mots-clés. Nous avons déjà vu qu'un mot-clé est un mot ou une expression, choisi généralement dans le titre ou le texte d'un document pour en caractériser le contenu et en permettre la recherche. Il constitue un point d'accès. Il est à distinguer d'un descripteur, qui est un terme normalisé dans un thésaurus ([1] Boulogne). Un descripteur est, en effet, un terme retenu dans un thésaurus pour représenter sans ambiguïté une notion contenue dans un document ou dans une requête de recherche documentaire. Ce peut être un nom commun ou un nom propre (nom géographique, de société, de personne, etc.), une locution, un mot composé ou un groupe de mots ([1] Boulogne). Ainsi, un mot-clé est issu du langage naturel et est dégagé par l'analyse documentaire. Un descripteur provient d'un langage contrôlé et normalisé. L'indexation revient, en quelque sorte, soit à utiliser les mots-clés retenus au cours de l'analyse documentaire (c'est ce que nous avons appelé plus haut l'indexation libre), soit à les traduire dans un langage documentaire, à les transformer en descripteur (il s'agit de l'indexation contrôlée).

1.4 Indexation automatique

Il est possible d'avoir recours à l'indexation automatique, c'est-à-dire une indexation effectuée exclusivement par des moyens informatiques. Elle peut prendre différentes formes : indexation en texte intégral (index constitué de tous les mots du document) ; indexation assistée par ordinateur (repérage de concept significatifs du document pour les caractériser par des descripteurs validés ou non par un professionnel) ; indexation par assignation (sélection des termes significatifs sur la base d'un langage documentaire) ([9] Raïs, p. 5). Pour ce faire, il est nécessaire d'utiliser des outils de traitement automatique de l'information et, en particulier, de l'information textuelle qui fait appel aux possibilités offertes par l'analyse linguistique automatique par l'indexation de l'information pertinente et/ou la production de métadonnées descriptives ([9] Raïs, p. 5) [9].

9. Il faut signaler quelques notions, assez récentes, qui associent indexation et nouvelles technologies. C'est le cas, par exemple, de « tag » (que l'on peut définir comme une marque

2 La recherche d'informations

Les langages documentaires jouent également un rôle dans la recherche d'informations. Essayons de voir comment ils interviennent à cette étape et en quoi ils peuvent être utiles.

2.1 Définition

En documentation, on distingue volontairement la « recherche d'informations » de la « recherche de l'information ». La première renvoie à un ensemble de méthodes, procédures et techniques permettant, en fonction de critères de recherche propres à l'usager, de sélectionner l'information dans un ou plusieurs fonds de documents plus ou moins structurés ([1] Boulogne). La seconde, qui découle donc de cette première définition, est un ensemble de méthodes, procédures et techniques ayant pour objet d'extraire d'un document ou d'un ensemble de documents les informations pertinentes ([1] Boulogne).

On peut considérer que la première parle des étapes pour trouver l'information, la seconde évoque le type d'informations recherchées.

2.2 Méthodologie

Toute recherche d'informations suppose trois phases successives ([1] Boulogne) :

a) une recherche bibliographique des références de documents pertinents ;

b) une recherche documentaire, c'est-à-dire une recherche bibliographique complétée par la recherche (l'acquisition) des documents eux-mêmes. J'y reviendrai plus loin ;

c) une recherche de l'information, qui n'est autre que le repérage de l'information dans les documents sélectionnés.

D'après Rais ([9] Raïs, p. 5), « la recherche d'information, manuelle ou informatisée, part de l'expression du besoin de l'utilisateur et consiste à fournir, en fonction d'une demande définie et spécifique, les éléments d'information correspondants sélectionnés dans un ou plusieurs fonds de documents structurés ou non ».

---

ou une balise destinée à l'identification, la description ou la mise en forme d'une élément de document), de « nuage de mots-clés » (ou tag clouds qui correspondent à une représentation visuelle des mots-clés les plus utilisés sur un site. C'est une sorte de condensé sémantique d'un document dans lequel les concepts sont associés à une unité de taille permettant de faire ressortir leur importance sur un site) ou encore de « folkosonomie » (néologisme désignant un système de classification collaborative décentralisée spontanée, fondé sur une indexation effectuée par des non-spécialistes).

Toutes les modalités de recherche mises en place ont pour but d'améliorer l'efficacité, l'exhaustivité et la pertinence de la recherche d'informations ([9] Raïs, p. 5). L'utilisateur doit pouvoir interroger toutes sources et/ou ressources d'informations de manière transparente. Il doit également pouvoir trouver ce qu'il cherche en limitant au maximum les risques de bruit (c'est-à-dire les réponses obtenues mais non pertinentes par rapport à la question) et de silence (réponses pertinentes existant dans le fonds interrogé mais non obtenus au moment de la recherche).

Le bruit est facilement repérable, ce n'est pas le cas du silence. On utilise pour cela des indicateurs tels que le taux de précision, qui est le ratio entre le nombre de documents pertinents trouvés et le nombre total de documents trouvés, et le taux de rappel, qui est aussi un ratio, cette fois entre le nombre de documents pertinents trouvés et le nombre de documents pertinents existant dans le fonds interrogé.

La recherche d'informations peut être libre (en langage libre sur le titre, le résumé et le texte) ou contrôlée (grâce aux descripteurs utilisés au moment de l'indexation).

2.3 Typologie

On peut considérer trois grands types de recherche qui correspondent en fait aux types d'informations recherchées.

2.3.1 La recherche factuelle

Elle correspond à un besoin de renseignements ponctuels. Il s'agit d'informations factuelles comme des données précises, des faits (par exemple, l'année de naissance d'un auteur…), des chiffres… Ces informations correspondent, en fait, à des « entités nommées » comme le nom des personnes, un organisme, une date… dans les métadonnées.

Auparavant, ces recherches étaient effectuées grâce à des annuaires, calendriers, dictionnaires…. Aujourd'hui, de nombreux répertoires ont été informatisés. Leur version informatique est assez performante en raison de la simplicité de la structure des notices, des relations entre les champs interrogeables et les champs de réponses, ce qui réduit les risques d'ambiguïtés. Les champs servant à l'interrogation sont facilement identifiables par l'utilisateur et sont logiques, voire intuitifs (numérique pour un code postal, textuel pour une profession). Pour les champs textuels, leur interrogation est facilitée grâce au recours à des concepts courants et aux renvois.

2.3.2 La recherche courante

Elle correspond à un besoin de renseignements liés à un contexte ou un domaine précis. Elle permet de trouver des informations de mises à jour sur ce domaine. Elle est souvent effectuée sur la littérature de l'année en cours et renvoie donc à l'actualité du domaine en question. C'est pour ce type de recherche que sont utilisés les fils RSS.

2.3.3 La recherche documentaire

Ce type de recherche renvoie à l'étude approfondie d'une question pour la réalisation d'une synthèse, d'un rapport… Ce sont des activités qui demandent la sélection d'un corpus de documents fournissant les connaissances nécessaires sur une question. Les informations recherchées sont donc de plusieurs types, mais elles ont pour but de couvrir une question.

2.4 Recherche d'informations et moteur de recherche

La recherche d'informations se fait de moins en moins par un spécialiste du domaine ou par un professionnel de l'information et de la documentation. Elle se fait de plus en plus directement par l'utilisateur final. Aussi les moteurs de recherche doivent proposer des interfaces d'accès à l'information qui soient simples et universels ([9] Rais, p. 24).

D'après Rais ([9] Raïs, p. 8), la recherche documentaire repose sur l'utilisation des outils de recherche et d'analyse de l'information textuelle pour améliorer l'efficacité de l'identification des documents pertinents dans les bases de données documentaires.

Les moteurs de recherche rendent, en général, possible l'utilisation d'une syntaxe libre, permettent également de bénéficier d'une extension automatique de la requête aux notions proches pour éviter le silence, de pouvoir s'appuyer sur un référentiel terminologique (quand il existe), d'accéder à une information dans d'autres langues que celle de la requête, enfin de bénéficier d'une aide à la lecture des résultats ([9] Raïs, p. 8).

Il faut bien distinguer l'interrogation en texte intégral de l'interrogation en langage naturel. Pour la première, le principe repose sur le fait que la question posée par l'utilisateur subit le même traitement que les documents de base (soit une indexation automatique). L'utilisateur doit formuler sa requête en respectant précisément les règles de syntaxe du langage de requête (termes de la recherche combinés avec des opérateurs booléens [10] ou de proximité [11]).

10. Les opérateurs booléens sont ET/AND (utilisé pour que les termes d'une recherche soient contenus dans les résultats), OU/OR (pour que l'un ou l'autre des termes d'une

Pour la deuxième, l'utilisateur peut formuler sa question dans la forme la plus naturelle possible, sans contrainte terminologique. C'est le système qui analyse et reformule automatiquement la requête dans la forme reconnue par le système. « L'usage des techniques de recherche en langage naturel est lié au développement des portails, point d'entrée unique pour un usager à un ensemble de ressources internes et externes, et à la nécessité de proposer des accès simples et conviviaux, indépendant du mode de gestion des bases de données par derrière qui doit être transparent pour l'utilisateur » ([9] Raïs, p. 8) [12].

recherche soient contenu dans les résultats), SAUF/AND NOT (pour que le terme suivant cet opérateur ne figure pas dans les réponses).

11. L'opérateur de proximité est NEAR. Il est présent sur quelques moteurs et peut s'avérer très pratique pour rechercher des termes qui ont un point commun, mais qui dans le texte sont séparés par d'autres mots. La proximité désignée par cet opérateur varie en fonction des moteurs de recherche. Certains permettent de spécifier le nombre de termes compris entre les mots de la requête.

12. Ce type d'interrogation est, par exemple, souvent utilisé dans le cadre d'entreprises de téléphonie. Pour plus d'informations, on pourra se reporter au mémoire de Desfriches Doria ([3] Desfriches Doria).

3 La valorisation de l'information

Par valorisation, il faut entendre accessibilité, visibilité et diffusion de l'information et/ou du document. Pour une revue scientifique, ces points sont essentiels. Dans tous les cas, ils sont dépendants des éléments précédemment décrits en matière d'indexation et de recherche d'informations.

3.1 Quelques définitions

Par accessibilité, il faut entendre « accès au contenu ». Dans un système d'information tel qu'Arlette Boulogne ([1] Boulogne) le définit :

> Ensemble d'éléments en interaction et formant un tout organisé et cohérent, mis en œuvre pour gérer, stocker et permettre l'accès à l'information et définis tant au niveau des politiques que des procédures et des ressources matérielles et humaines. En fonction du public ciblé, de leurs usages et des types de données gérées, ces systèmes possèdent des fonctionnalités spécifiques et parfois des intitulés particuliers [13].

Qui plus est, dans un système d'information documentaire, soit un système d'information assurant la gestion et la communication des informations utiles contenues dans des documents [14], il s'agit du mode de recherche d'informations proposé à l'utilisateur, lui permettant d'une part de repérer, de trier, d'autre part d'atteindre l'information ou le document recherché.

Ce mode de recherche dépend en grande partie de la façon dont le document a été traité en amont. En effet, la recherche peut prendre différents aspects. Cela peut être une recherche par mots-clés, à partir d'une liste contrôlée et utilisée pour représenter le contenu, ou une recherche plein texte ou en texte intégral, donc à partir de mots figurant dans les parties textuelles du document ou encore une recherche en langage naturel, à partir de mots relevant du langage naturel et associés à des traitements linguistiques et/ou sémantiques. Il est également possible de proposer à l'utilisateur une sélection de termes faisant partie ou non du document comme un plan de classement ([1] Boulogne).

---

13. Système d'information géographique (SIG), système interactif d'aide à la décision (SIAD), système d'information comptable, système d'information documentaire, système (d'information et) de gestion de documents techniques (SGDT), etc. Aujourd'hui ce terme de système d'information renvoie systématiquement à un dispositif informatisé.
14. Cette gestion implique le repérage et la validation des sources et ressources d'information, la collecte, le traitement, le stockage, la diffusion ainsi que les moyens de classement et de récupération des documents porteurs d'information en utilisant des outils classiques ou informatisés ([1] Boulogne).

Ce qu'il faut retenir de tout cela c'est qu'après avoir examiné les usages des utilisateurs, il est possible de définir certains de leurs besoins et, donc, de les aider dans leur recherche d'informations, en les orientant, en leur donnant des consignes et des conseils d'utilisation. Pour cela, il faut déjà connaître l'existence d'un site, que celui-ci soit visible.

S'il était possible de parler de visibilité avec les outils traditionnels, il est évident que ce terme prend véritablement toute son ampleur avec le développement du web. La visibilité rejoint d'un certain point de vue l'accessibilité, tout en étant plus superficielle en ce sens qu'elle n'est pas nécessairement liée au document, mais plutôt au site qui le contient. Elle se rapproche par conséquent de la notion de référencement :

> Ensemble d'opérations visant à favoriser l'accessibilité d'un site web à travers les outils de recherche de l'Internet, en premier lieu les moteurs de recherche et les annuaires ou répertoires. Le référencement consiste à : a) organiser les données et leur attribuer des éléments descriptifs (métadonnées) afin d'aider le public cible à découvrir ou retrouver le site ; b) enregistrer le site dans les outils de recherche ; c) exercer une veille sur les effets de cet enregistrement. Ces opérations doivent tenir compte des particularité du public cible d'une part, des caractéristiques techniques des outils de recherches d'autre part. ([1] Boulogne)

En développant l'accessibilité et la visibilité de l'information, on pourrait se contenter de cela et on aurait déjà fait une grande part pour sa valorisation. En ajoutant la diffusion, on ouvre une autre porte qui devrait toucher encore plus d'utilisateurs et, pourquoi pas, rendre fidèles ceux déjà présents. Par diffusion, il faut entendre la fourniture d'information répondant à une demande précise. Nous allons y revenir tout de suite car, grâce à la diffusion, nous entrons dans le domaine des moyens à mettre en place pour valoriser l'information.

3.2 Méthode et moyens

Pour proposer une valorisation de l'information, il faut revenir sur des notions telles que produits documentaires et/ou services documentaires. En effet, c'est en déployant ce type de moyens qu'il sera possible de mettre en valeur un système d'information.

Les produits documentaires sont des documents secondaires ou tertiaires [15], conçus pour répondre à des besoins d'information. Ils peuvent prendre des formes diverses : bibliographie, bulletin bibliographique ou de liaison, bulletin de sommaires, dossier documentaire, dossier de presse, DSI, revue de presse, état de la question, synthèse documentaire… ([1] Boulogne)

15. D'après l'AFNOR, *Vocabulaire de la documentation* (1987), il s'agit d'un document présentant soit une synthèse de documents primaires, soit un regroupement de signalements ou d'analyse de documents secondaires (bibliographie de bibliographies).

Un service documentaire a pour vocation de répondre à des besoins d'information. Il a pour objet d'assurer le transfert et l'utilisation de l'information vers l'utilisateur par des moyens diversifiés : service question/réponse, mise à disposition d'espace de consultation, DSI, etc.

Ainsi, en tenant compte des besoins spécifiques des utilisateurs, on peut être amené à proposer par exemple une diffusion sélective d'information (DSI ou « profil »), une veille informationnelle et/ou des alertes.

Introduisons là aussi les définitions de ces différents produits et/ou services.

Une DSI correspond à la fourniture personnelle, systématique et périodique d'informations (et/ou de documents) qui répond, pendant un temps déterminé, à une demande d'information correspondant à un domaine d'intérêt (ou « profil »). Il existe des profils standards prédéfinis par les fournisseurs d'information ou des profils personnalisés construits à la demande précise d'un usager. Les logiciels documentaires permettent de sauvegarder une ou plusieurs équations de recherche, exécutables périodiquement. Les services dits « push » utilisent ces méthodes de diffusion sélective de l'information et offrent en plus à l'utilisateur la maîtrise de l'abonnement/désabonnement au service ainsi que le choix des sources et termes de recherche ([1] Boulogne).

Retenons donc que, pour modéliser ces profils d'intérêt personnalisé, on peut soit faire intervenir l'usager en lui demandant de sélectionner les termes et les sources qui l'intéressent, soit lui proposer des choix de termes et de sources préalablement établis en fonction de l'analyse des concepts exprimés dans un corpus documentaire sélectionné par l'utilisateur.

Une veille informationnelle est un dispositif organisé, intégré et finalisé de collecte, traitement, diffusion et exploitation de l'information qui vise à rendre une organisation quelle qu'elle soit, capable de réagir, à moyen et long termes, face à des évolutions ou des menaces de son environnement, que celles-ci soient technologiques, concurrentielles, sociales, etc. ([1] Boulogne).

Pour arriver à un tel dispositif, la mise en œuvre d'outils de traitement (automatique ou non) du contenu est nécessaire car ils vont permettre le repérage des signes avant-coureurs liés à un champ d'application dans le flux informationnel.

Par exemple, grâce aux outils de traitement automatique de la langue (ou TAL), il sera *idéalement* possible :

-de filtrer l'ensemble des documents proposés sur le web ou dans un système d'information pour repérer, sélectionner et extraire l'information pertinente ;

-de classer l'information dans des catégories en fonction des centres d'intérêt des utilisateurs.

Remarquons néanmoins que ces catégories sont créées *a priori* par l'être humain et non pas par les outils de TAL.

De plus, précisons qu'une classification automatique des connaissances a pour objectif une organisation des résultats par classes de documents en fonction de leur plus ou moins grande pertinence par rapport à la question posée.

La classification automatique peut être supervisée (regroupement des éléments dans des classes pré-définies d'un plan de classement) ou non (regroupement des éléments dans des classes non déterminées permettant la répartition des éléments dans des groupes qui évoluent au fur et à mesure des besoins et selon les recherches effectuées) ([9] Raïs, p. 9).

Il existe plusieurs façons de mettre en valeur l'information, soit en utilisant des moyens pour la rendre accessible et visible, soit en utilisant des moyens pour la diffuser en allant au-devant de l'utilisateur, soit en utilisant les deux, c'est d'ailleurs ce qui se fait généralement. Dans tous les cas, on constate que le traitement de l'information en amont est essentiel car c'est lui qui va déterminer sa mise en valeur ensuite. L'indexation semble, quoi qu'en disent certains, être toujours au cœur de la problématique de l'information et de sa valorisation.

Avec Nadia Raïs ([9] Raïs, p. 10), on peut considérer que le rôle des professionnels de l'information est central non seulement pour déterminer des critères de sélection et d'évaluation des sources d'informations pertinentes, pour fédérer des sources d'information sur un portail ou site intranet, mais aussi et surtout pour enrichir et structurer l'information (importance des métadonnées), pour définir des modalités d'accès et de recherche, pour afficher les résultats, pour déterminer des produits personnalisés…

4 Focal sur le web sémantique

Les langages documentaires à la fin des années 1990 ont été malmenés : on ne leur voyait pas d'avenir. Pourtant, depuis trois ou quatre ans, on se rend compte que la question de l'indexation des ressources électroniques est la question centrale de la nouvelle économie du savoir en émergence ([10] Serre). En effet, face à la profusion d'informations électroniques ainsi qu'à son hétérogénéité et au besoin d'aide pour les usagers afin de s'y retrouver, le défi majeur est celui de la représentation de l'information ([2] Chaudiron, p. 31).

Plusieurs constats à propos des informations disponibles sur le web peuvent être avancés :

- On peut facilement avoir accès au document sur le web, mais on est moins sûr de trouver l'information pertinente.

- On trouve toutes les formes et tous les modes d'indexation des ressources sur le web : des classifications (annuaires ou répertoires), des listes de mots-clés, du texte intégral, des métadonnées, des thésaurus…

- On oppose souvent l'indexation documentaire par un langage contrôlé à l'indexation automatisée sur le texte intégral, alors qu'il serait préférable de voir leurs complémentarités.

En suivant les propositions de Stéphane Chaudiron ([2] Chaudiron), on peut considérer trois voies pour représenter le contenu informationnel d'un document ou d'une collection de documents.

La première voie est liée à la conception proprement dite des langages documentaires qui ont permis de faire une indexation contrôlée.

La deuxième voie consiste à représenter les documents repérés par le crawler d'un moteur par les mots contenus dans ces documents. Cette indexation libre se fait par extraction automatique des unités du document qui peuvent être considérés comme des mots-clés.

La troisième voie est en relation avec la structuration des « sources d'information, notamment celles du web, en rendant explicites les relations sémantiques qui peuvent exister entre les différentes unités informationnelles que contiennent les documents » ([2] Chaudiron, p. 31). C'est dans cette voie que s'inscrivent les travaux sur le « web sémantique » et sur l'utilisation des métadonnées qui reposent en grande partie sur le langage XML et ses variantes.

Quoiqu'il en soit, ces trois approches de la représentation du contenu concourent à faciliter la recherche d'informations et peuvent être utilisées de manière complémentaire.

D'après Alexandre Serre ([10] Serre), on peut considérer quatre directions sur les travaux concernant le web sémantique : la première porte sur l'identification stable des concepts et des objets du web (donc sur les normes d'identification des documents électroniques [16]) ; la deuxième est sur le métalangage XML et la normalisation des modèles de structuration des documents ; la troisième aborde les outils et les normes d'indexation des documents, comme les métadonnées Dublin Core… ; enfin la dernière traite de la construction d'ontologies et de thésaurus interopérables.

Arrêtons-nous sur le XML (ou eXtensible Markup Language). Malgré la signification de son acronyme, le XML est plus un métalangage qu'un langage informatique. Il permet de concevoir des méthodes de description des documents très fines en distinguant la structure physique du document (succession de lignes, de paragraphes, de pages…) de sa structure logique (titre, résumé, intertitres…). Le XML ne remplace pas le HTML, mais l'enrichit. Un document enregistré en XML peut être publié sous d'autres formats d'édition (HTML, PDF…). Les différents systèmes de métadonnées [17] offrent une cohérence dans la description des documents, ce qui permet l'interopérabilité des données.

D'après le *Grand dictionnaire terminologique* de l'Office québécois de la langue française [18], l'interopérabilité correspond à la capacité que possèdent des systèmes informatiques hétérogènes à fonctionner conjointement, grâce à l'utilisation de langages et de protocoles communs, et à donner accès à leurs ressources de façon réciproque. Elle implique qu'un programme tournant sur un système ouvert fonctionnera également sur un autre système. L'interopérabilité définie ici est l'interopérabilité technique, mais il en existe d'autres types dont l'interopérabilité sémantique qui est associée à un mode de description de l'information contenue dans une base de données (cette description forme les métadonnées). L'interopérabilité a besoin de plus qu'une bonne connectivité technique puisqu'elle nécessite

16. Il faut ici mentionner le DOI (Digital Object Identifier) qui est le numéro international identifiant et localisant un objet documentaire publié sur Internet. Il est composé d'un préfixe identifiant l'éditeur et d'un suffixe identifiant l'objet chez l'éditeur, séparés un slash. Le dispositif DOI, sous la responsabilité de la Fondation internationale DOI, comporte trois éléments : un identifiant pour les objets, un répertoire et une base de données d'informations sur les objets gérés par chaque éditeur. Le répertoire est un système central unique de résolution des identifiants, les traduisant dans les URL des systèmes des détenteurs de droit (Norme ISO Z39.84, 2000).
17. Se reporter au livre de l'ADBS sur les métadonnées ([16] Calderan).
18. Voir http://www.granddictionnaire.com/BTML/FRA/r_Motclef/index800_1.asp (consulté le 16 septembre 2009).

l'utilisation d'éléments comme des interfaces de programmation et des formats de données standardisés. L'interopérabilité peut être totale ou partielle.

Retenons que l'interopérabilité a connu un développement important dans le monde de la recherche scientifique, qui plus est dans le domaine l'édition électronique scientifique. En effet, d'après Stéphane Pouyllau [19], elle s'est développée autour du protocole OAI-PMH qui expose les métadonnées sous la forme d'un flux XML et selon le schéma Dublin Core. Grâce à ce protocole, il est possible de construire des portails thématiques adossés, par exemple, à une thématique de revue et de mettre en place un système de moissonage des métadonnées dans le but de faciliter la recherche et la diffusion d'informations. Il suffira de regarder le site de Revues.org, qui tend vers cet objectif, pour mieux s'en rendre compte.

Finalement, il apparaît que tout tourne autour de l'idée de structuration des informations et/ou de données, une structuration à la fois sous la forme de métadonnées et sous la forme d'une indexation libre et/ou contrôlée. On peut donc imaginer qu'idéalement, plus l'information sera structurée, plus elle sera accessible, plus les réponses aux requêtes seront pertinentes et, sans doute, sera-t-il plus facile de mettre sur pied des produits et/ou services documentaires répondant mieux aux besoins des utilisateurs.

19. Voir http://www.digitalhumanities.cnrs.fr/wikis/edelec-shs/index.php/L%E2%80%99interop%C3%A9rabilit%C3%A9_:_gestion_et_circulation_de_l'information_scientifique (consulté le 09 octobre 2009).

Deuxième partie :
Revues en sciences humaines et sociales, mots-clés et recherche d'informations. Étude de cas

Dans cette partie, je vais rendre compte d'une enquête menée auprès d'auteurs (chercheurs, enseignants-chercheurs, étudiants…), rédacteurs et secrétaires de rédaction au sujet des mots-clés dans des revues en sciences humaines et sociales. Cette enquête a pour origine un questionnement sur ces mots-clés présents dans la plupart des revues en SHS, voire dans la plupart des revues scientifiques en général. En effet, à partir des années 1970, il a été demandé aux auteurs d'article de fournir, en plus de leur texte, un résumé et 5 à 8 mots-clés, parfois aussi de les traduire en anglais et/ou dans une autre langue comme l'allemand et l'espagnol. Ainsi, depuis plusieurs années, on demande à des spécialistes de donner des mots-clés spécialisés. Les revues peuvent donc être perçues comme des réservoirs à mots-clés spécialisés. Mais qu'en fait-on ? comment s'en sert-on ?

Je tenterai de répondre à ces questions en dressant une sorte d'état des lieux de la situation dans quelques revues du pôle éditorial. Ma référence principale en matière de revue scientifique sera la revue pour laquelle je travaille, soit le JSA *(Journal de la Société des Américanistes)*. Aussi certains aspects décrits resteront à un niveau très général car il sera difficile de prendre en compte toutes les variations caractérisant chacune des revues.

Je propose de décrire, dans une première partie, les étapes de production et l'organisation de l'information dans une revue. Cela permettra de poser le décor. Même si, d'une revue à l'autre, l'organisation de l'information peut prendre des formes différentes, les étapes de la chaîne de production sont à peu près les mêmes et l'information est, d'une façon ou d'une autre, organisée. Dans une deuxième partie, je m'interrogerai, d'une part, sur l'élaboration de ces mots-clés et sur ce qu'ils représentent pour les personnes qui les fabriquent, d'autre part, sur la façon dont ils sont perçus et manipulés par les secrétaires de rédaction, rédacteurs et directeurs de revues. Enfin, la dernière partie rendra compte de quelques aspects de la recherche d'informations mise en place par les auteurs rencontrés. Il y sera question de méthodes de recherche et de typologie des informations recherchées, puis des outils nouveaux et anciens mis à leur disposition dans les revues.

Retenons que les auteurs de texte dans une revue sont aussi les lecteurs de cette même revue. Ainsi les personnes intéressées par le domaine d'une revue, donc les utilisateurs, peuvent également être les personnes qui fournissent la matière, soit les producteurs. Aussi, le public visé ou touché par ces problématiques est, dans le cas du pôle éditorial de la MAE, spécialisé dans les domaines de l'ethnologie, de l'archéologie et de la sociologie. Il s'agit en effet de chercheurs, d'enseignants-chercheurs, d'étudiants en doctorat, en master 1 et 2. Il faudra, peut-être parfois, distinguer le chercheur ou enseignant naviguant dans son domaine de recherche depuis des années, de l'étudiant débutant ou pas car si l'un possède une grille de lecture et sait facilement reconnaître un document ou une information valide et pertinente, l'autre n'a pas encore acquis tous les outils nécessaires à cette sélection.

5 Une revue en SHS : un système d'informations organisées ?

Un pôle éditorial, tel que celui de la Maison René-Ginouvès, se compose de plusieurs revues aux statuts, domaines et sujets différents. Une revue se compose de textes variant dans leur forme et leur contenu selon la rubrique pour laquelle ils sont écrits. Un texte se compose d'un titre, d'un éventuel sous-titre, d'un (ou plusieurs) auteur(s), d'intertitres, de paragraphes retraçant le cheminement de la pensée et l'enchaînement des idées, des références citées, d'un résumé (en plusieurs langues) et de 5 à 8 mots-clés (en plusieurs langues également, mêmes langues que pour le résumé). Un paragraphe se compose de phrases, de mots. De plus, des données brutes (telles que des chiffres, des chants…) peuvent être insérées dans le (ou jointes au) texte. Le plus petit élément commun dans un texte, c'est le mot, mais le mot n'a de sens que pris dans une phrase et la phrase n'a de sens que dans un paragraphe, le paragraphe ne prend toute son ampleur qu'avec les autres paragraphes, c'est ainsi que se construit un texte.

Dans un esprit assez similaire, ne peut-on envisager une revue scientifique comme un système d'informations organisées et/ou structurées, reconnu dans le milieu scientifique correspondant au domaine qu'elle couvre ? Il est en effet possible de dire qu'elle propose en quelque sorte une information structurée que l'on peut classer et catégoriser. Pour renforcer cette idée de façon plus explicite, disons qu'un texte publié dans une revue scientifique peut appartenir à telle ou telle rubrique, donc être repérable par la catégorie de textes à laquelle il appartient (un article, un compte rendu de lecture, une nécrologie…). De plus, il peut être classé par domaine selon les thèmes qu'il traite : par exemple, dans le JSA, il peut s'agir d'un texte en ethnologie, archéologie, ethnolinguistique, ethnohistoire… Cette façon d'organiser les textes donne d'emblée des points de repère non négligeables pour trouver l'information recherchée. Allons plus loin et entrons dans la revue proprement dite.

5.1 Quelques repères sur l'organisation des informations dans une revue scientifique

Une revue scientifique est, à la fois, un objet matériel et/ou virtuel [20] et un rassemblement de spécialistes d'un domaine. L'objet « revue » est composé, extérieurement, d'un titre,

20. Je fais ici référence aux revues uniquement électroniques. Au sein de la MAE, il en existe deux, mais elles ne font pas partie du pôle éditorial : il s'agit des Ateliers du LESC (http://ateliers.revues.org) et d'Arkeotek (http://www.arkeotek.org/).

souvent représentatif du domaine sur lequel elle est spécialisée et d'une périodicité marquée de différente façon (par l'année, la tomaison ou le volume, le numéro…). Dans ses pages intérieures, une revue est divisée en rubriques. Par exemple, dans le JSA, on trouve comme rubriques : articles, notes de recherche, comptes rendus d'ouvrages, chronique, nécrologie… Chaque rubrique a un objectif scientifique clairement défini dans la politique éditoriale de la revue ; elle peut, parfois, avoir un responsable désigné par le comité de rédaction et chargé de trouver des textes.

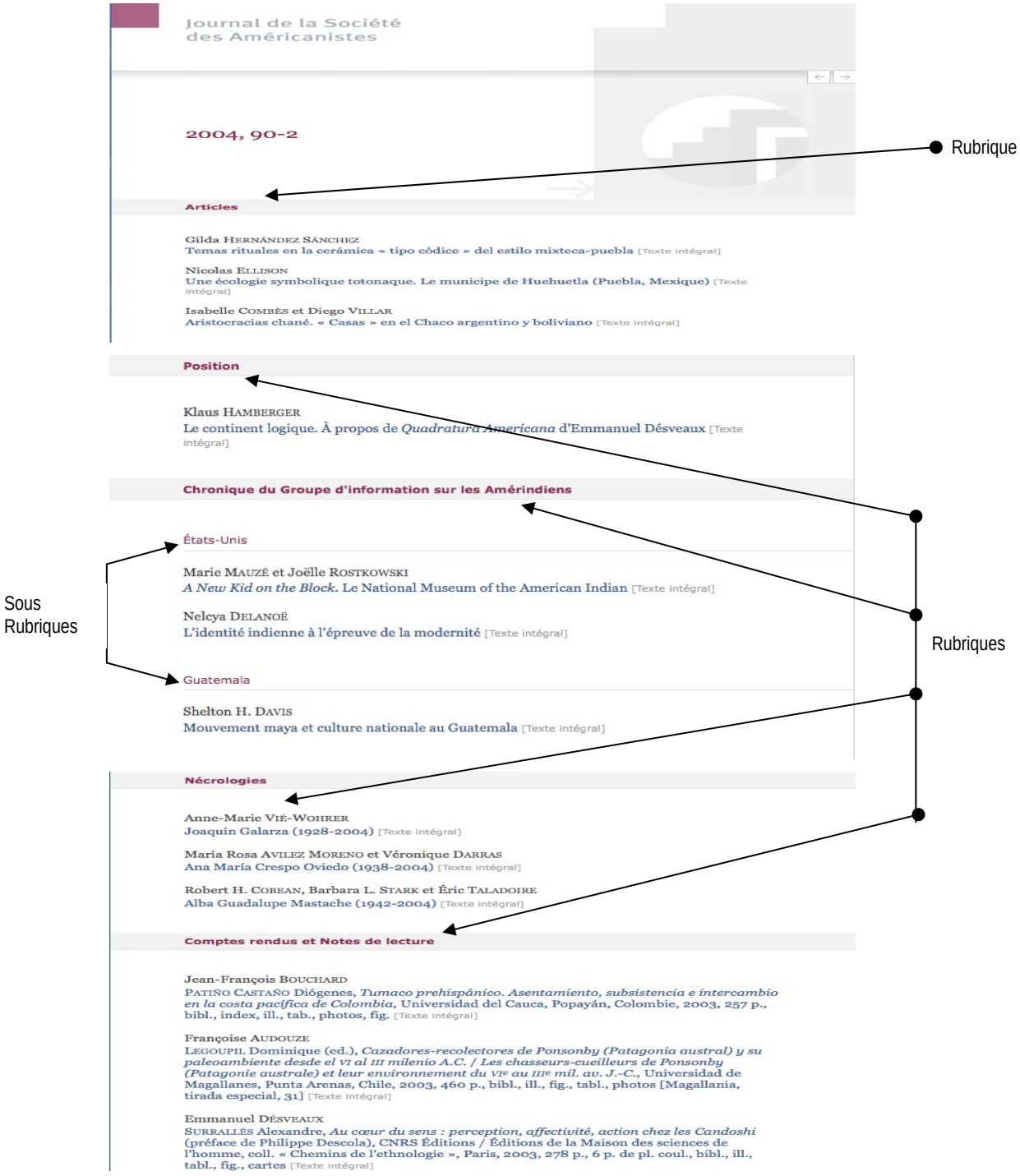


Fig. 2 – Copies d'écran des rubriques du JSA.

Une revue scientifique, c'est aussi un rassemblement de spécialistes du domaine. Ils sont réunis sous la forme de comité de rédaction et/ou de conseil scientifique. Le comité a un rôle décisionnaire à plusieurs niveaux. À propos des textes soumis, il décide des experts qui vont l'évaluer et du sort qui va leur être imposé. À propos de la revue, il décide de la répartition des textes, c'est-à-dire du sommaire, selon chaque volume, de la politique éditoriale et de la politique de diffusion de la revue. Le conseil scientifique rassemble des spécialistes selon les sous-sections du domaine et permet de rassembler des experts potentiels. Il peut arriver que conseil et comité soient confondus.

Bien évidemment, une revue n'aurait aucun sens sans les textes qui la composent. Dans un texte, on retrouve toujours le même type d'information, même si le contenu à proprement parler n'a rien à voir d'un texte à l'autre. Le contenu d'un texte peut être décomposé en parties, sous parties, paragraphes, phrases, mots. Il est parfois accompagné d'illustrations (figures, photographies, tableaux, schémas…). Généralement, certaines informations sont repérables très facilement : c'est le cas du titre, de l'éventuel sous-titre, du nom et prénom de l'auteur, de son affiliation institutionnelle, du résumé, des mots-clés, des intertitres, des notes de bas de page ou de fin de texte, des références citées. Elles sont repérables facilement car toujours positionnées au même endroit et présentées de la même façon.

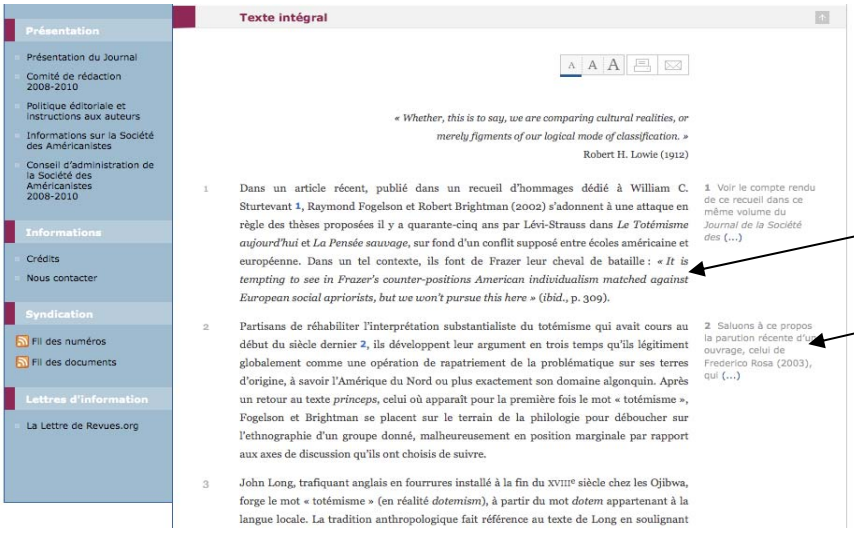
Ces informations sont aussi bien repérables dans la version papier que dans la version électronique.





Résumé

Mots-clés



Texte

Notes qui, dans la version papier, se trouvent en fin de texte



Références citées

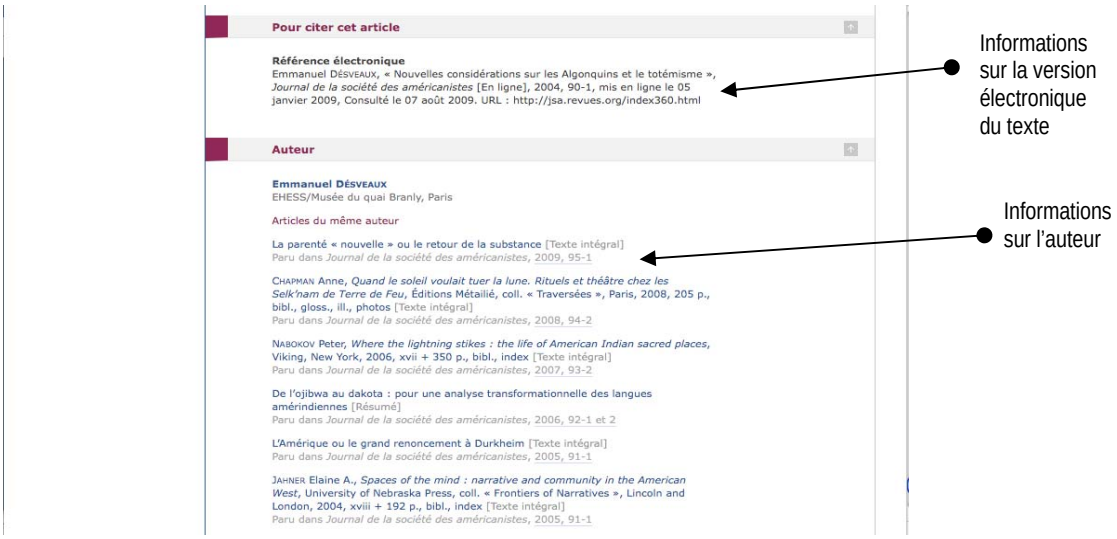


Fig. 3 – Copies d'écran d'une page de présentation d'un texte dans le JSA.

5.1.1 Étape de production d'un document dans sa version papier et dans sa version électronique

Les étapes pour produire un texte qui sera publié dans une revue se divisent *grosso modo* en trois grands moments : évaluation du texte ; préparation de la copie ; publication. Je me suis appuyée sur les étapes mises en place pour la fabrication du JSA.

5.1.2 Évaluation du texte reçu

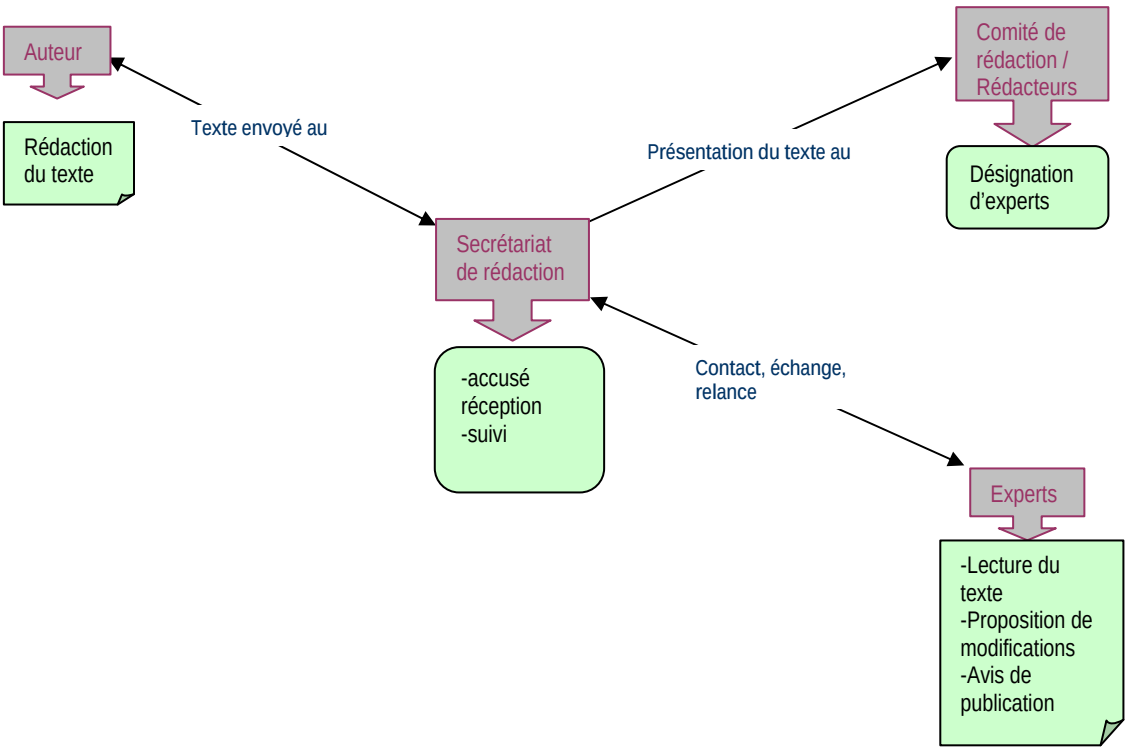


Fig. 4 – Schéma acteurs/activités au moment de l'évaluation d'un texte.

Cette étape est importante car c'est au cours de celle-ci que le sort du texte va être décidé. Les experts vont émettre un avis. Ils peuvent accepter le texte pour publication directement. Le plus souvent, ils proposent des modifications plus ou moins nombreuses et profondes. Enfin ils peuvent le refuser sans autre forme de procès. Le refus ne donne pas lieu à des suites, l'auteur doit trouver une autre revue et/ou réécrire entièrement son texte. Quand le texte est accepté, une fois que les demandes de modifications ont bien été prises en compte, il fait l'objet d'une préparation éditoriale, étape la plus longue du processus.

5.1.3 Préparation éditoriale de la copie

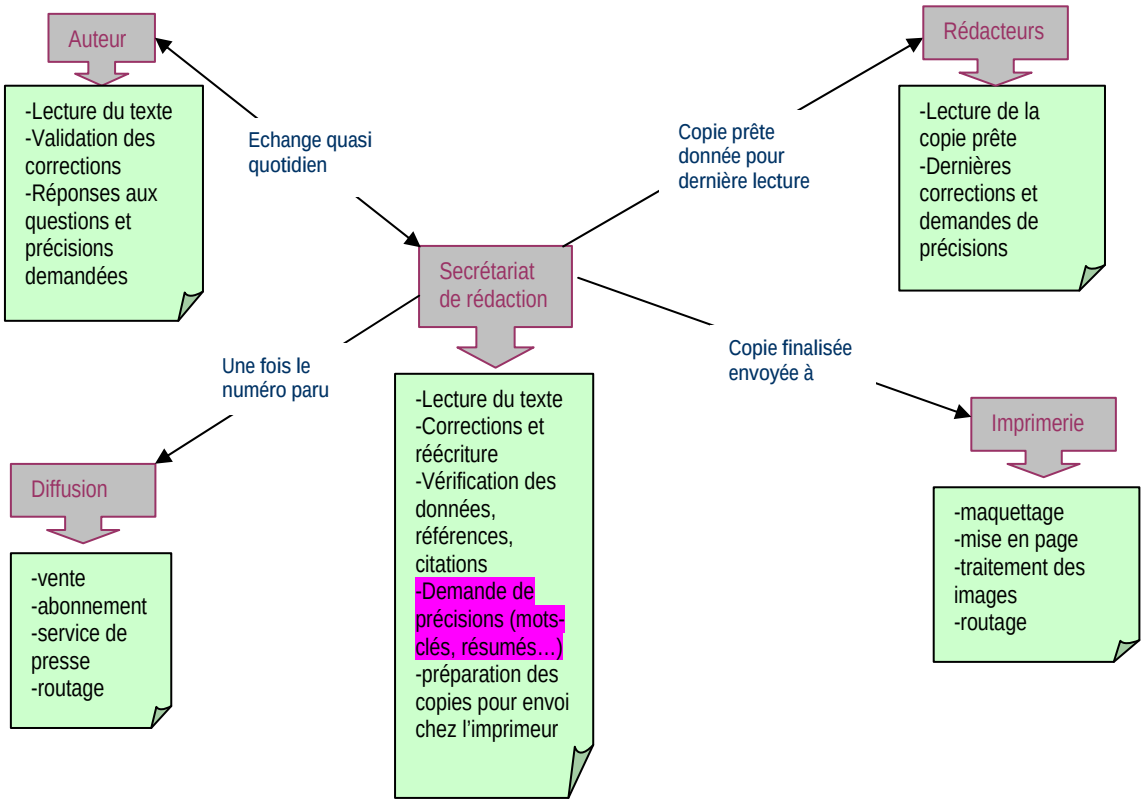


Fig. 5 – Schéma acteurs/activités au moment de la préparation éditoriale d'un texte.

Cette phase correspond à la préparation des textes avant l'envoi chez l'imprimeur. Elle peut être composée d'un grand nombre d'aller-retour entre le secrétariat de rédaction et les auteurs et a pour but d'envoyer à l'imprimeur un texte propre. C'est à ce moment-là que sont vérifiés ou demandés les mots-clés et résumé.

5.1.4 Préparation pour la publication

Cette dernière étape renvoie à la lecture des premières épreuves, des secondes épreuves et à l'envoi du bon à tirer (ou BAT) à l'imprimeur. Pour le JSA, les premières épreuves sont relues par les auteurs, le secrétariat de rédaction et parfois un rédacteur. Les auteurs envoient leur BAT en même temps que leurs corrections sur épreuves. Les secondes épreuves ne sont relues généralement que par le secrétariat de rédaction. Les pages portant des corrections sont envoyées à l'imprimeur, avec le BAT, cette fois, pour l'ensemble du numéro.

5.1.5 Mise en ligne dans le cas du JSA

Dans le cas du JSA, une fois les secondes épreuves envoyées à l'imprimeur, commence la préparation pour la mise en ligne du numéro. Toutes les corrections faites sur épreuves sont entrées dans la version Word des textes. Puis ceux-ci sont copiés et collés dans une feuille de style Lodel, comprenant des métadonnées Dublin Core. Enfin la feuille de style ainsi préparée est téléchargée sur la plateforme réservée à la revue et le texte est placé dans la rubrique correspondant.

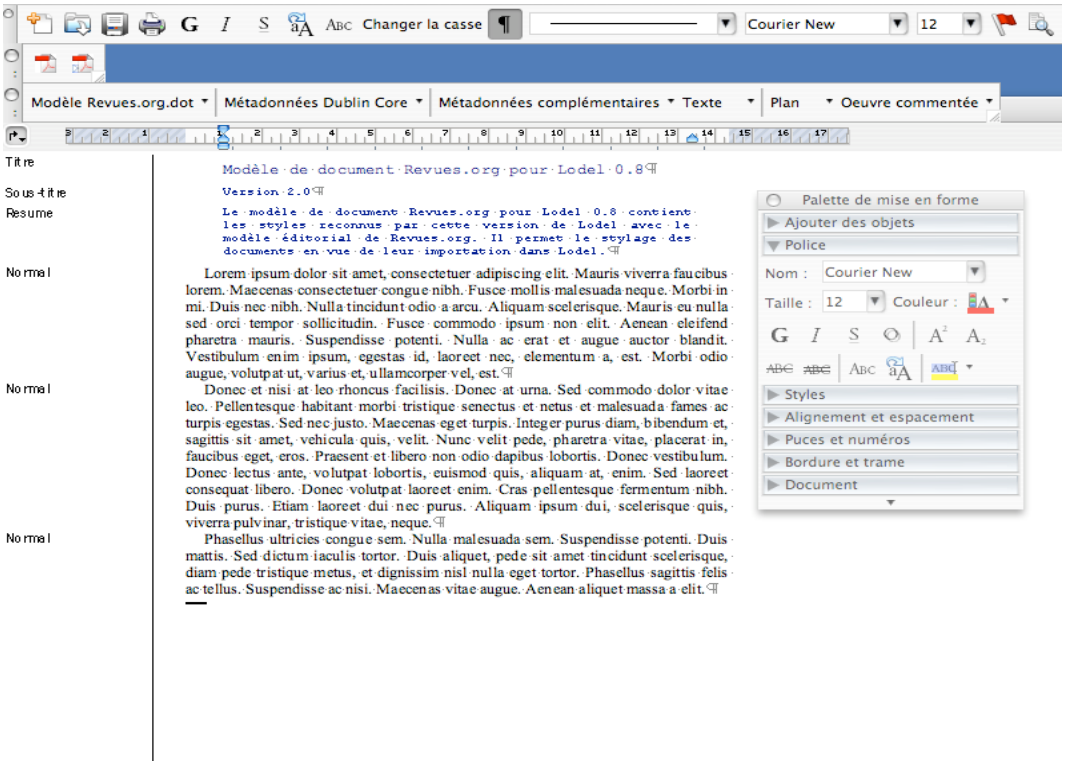


Fig. 6 – Copie d'écran du modèle de document Revues.org pour Lodel.

Les métadonnées de ce document se décomposent de la manière suivante :

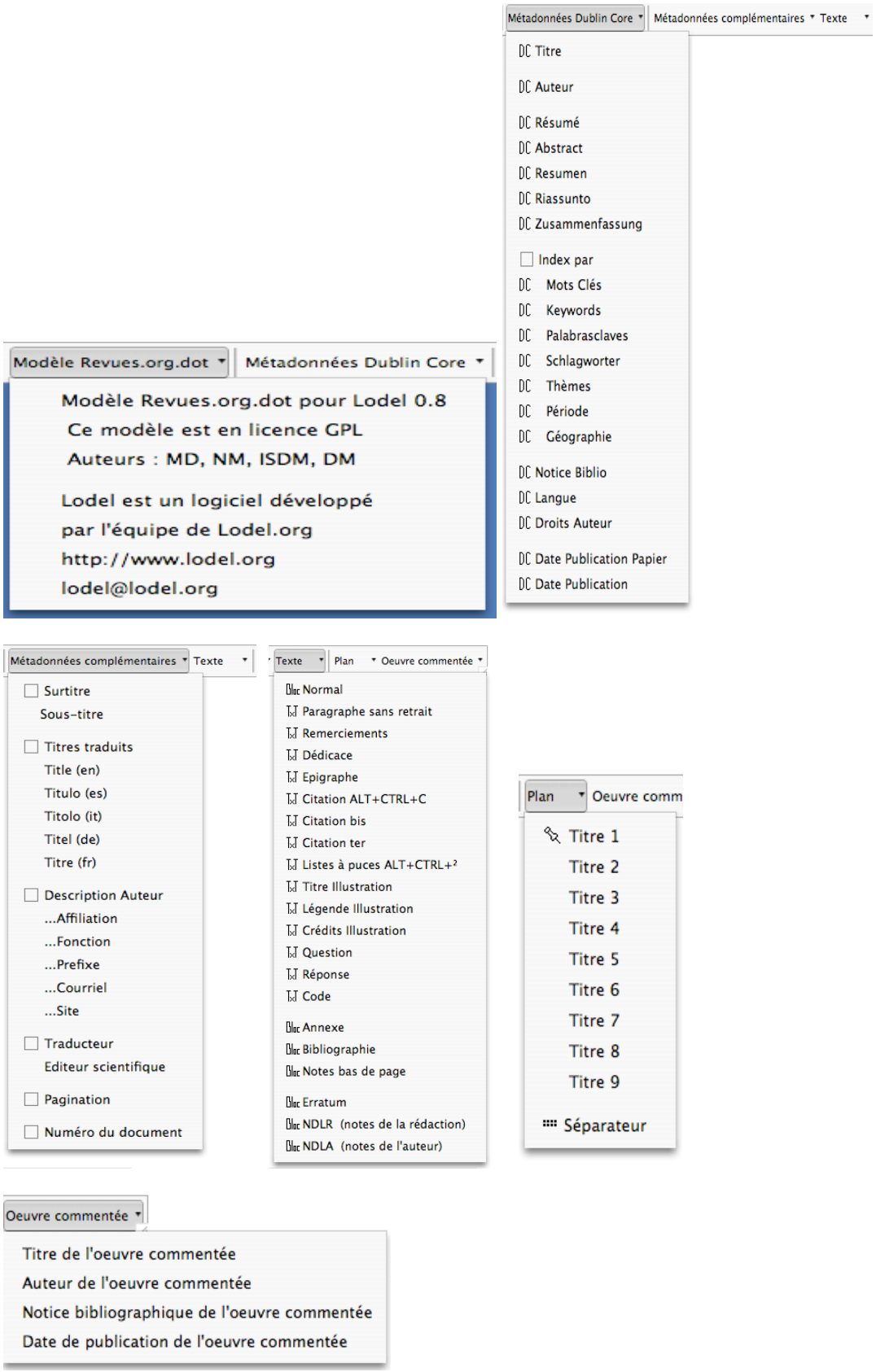


Fig. 7 – Copies d'écran des listes de métadonnées et de styles développées sur le modèle de document Revues.org pour Lodel.

Grâce à cette préparation sur la feuille de style et à l'utilisation de ces métadonnées, les informations concernant le texte sont structurées. Le contenu du texte l'est également, mais d'une façon plus légère. On peut, par exemple, retrouver facilement les citations, les notes, les références bibliographiques. En effet, chacune de ces parties est décrite grâce soit à l'application d'une métadonnée, soit à l'attribution d'un style. En revanche, les thèmes et/ou sujets abordés ne peuvent être mis en valeur que grâce à une indexation. Il existe, parmi les métadonnées Dublin Core retenues dans le modèle de document de Revues.org pour Lodel, une entrée correspondant à l'idée d'un index [21]. Celui-ci serait constitué à partir des mots-clés du texte et serait décliné en mots-clés, thèmes, période, géographie. Il est donc possible de penser que ces informations sur le contenu sont (ou peuvent être) organisées. Pour le JSA (je le montrerai plus loin), ce sont les mots-clés fournis par les auteurs qui aident à remplir ces entrées, donc à faire l'indexation des textes. Or ces mots-clés ne subissent aucun contrôle puisqu'ils sont conservés tels qu'ils ont été donnés par les auteurs. Intéressons-nous donc de plus près à ces mots-clés.

21. Dans le chapitre suivant, des informations complémentaires seront apportées sur la façon d'accéder à ces index et de visualiser les termes utilisés.

6 Mots-clés et résumés dans des revues en SHS

Comme je l'ai déjà dit, dans une revue scientifique, chaque article est accompagné de mots-clés et d'un résumé, traduits en plusieurs autres langues. Ces mots-clés et résumé sont fournis par le ou les auteurs du texte. On peut donc supposer qu'ils sont spécialisés puisqu'ils sont donnés par des spécialistes. C'est dans le courant des années 1970 que les revues scientifiques ont commencé à demander à leurs auteurs des mots-clés. On peut avancer, sans prendre trop de risque, que cela s'est produit en même temps que le développement de l'informatique documentaire, des bases de données et des langages contrôlés.

En dépit de cette concordance temporelle – qui pourrait expliquer pourquoi on a pu penser, à un moment donné, qu'il était important que les auteurs d'articles donnent des mots-clés correspondant au contenu de leur texte –, il est difficile de voir à quoi ces mots-clés sont censés servir [22]. Peut-on penser que certains d'entre eux (nom de peuple, nom géographique, thème général) ont pu être utilisé, par les secrétariats de rédaction (secrétaire et rédacteur/directeur), pour faire des index papiers des numéros de la revue ? ont-ils été tous utiles pour ces types d'index ? ont-ils servi à indexer les articles dans des bases de données ? ou ont-ils juste été une inspiration pour orienter et choisir les descripteurs pour une indexation fine ? comment les auteurs conçoivent-ils ces mots-clés ? quelles questions se posent-ils pour les fabriquer ? Voyons ce que les auteurs et les personnes travaillant dans les rédactions de revues en disent…

6.1 Enquête et questionnaire : brève présentation

Pour trouver des réponses à ces questionnements, j'ai mené une enquête auprès d'auteurs (chercheurs, enseignants-chercheurs, étudiants…) de texte pour savoir comment ils élaborent leurs mots-clés, auprès des secrétaires de rédaction et rédacteurs et/ou directeurs de revues pour savoir quelle attention ils portent à ces mots. Auprès de ces deux groupes de personnes, j'ai aussi essayé de voir comment ils se représentent l'utilisation de ces mots.

Sept auteurs ont bien voulu répondre au questionnaire [23]. Celui-ci a été composé en deux parties : la première aborde les méthodes mises en place pour rechercher des informations ; la seconde traite les mots-clés et le résumé. Pour rendre compte des données recueillies

22. On a pourtant montré plus haut que, pour le JSA, ils servent à constituer les index pour l'édition électronique. Or cette pratique est assez récente et très peu satisfaisante comme nous le verrons plus loin.
23. Voir annexe 5.

dans cette partie, j'inverserai les thèmes et je commencerai par les mots-clés, puis les méthodes de recherche d'informations.

Parmi ces sept auteurs, je n'ai pu m'entretenir qu'avec trois d'entre eux, les quatre autres ont seulement répondu au questionnaire sans faire d'entretien. Cinq sont chargés de recherche au CNRS (trois sont ethnologues ; deux sont archéologues) et les deux autres sont docteurs en ethnologie (c'est-à-dire sans poste pour le moment). Deux, seulement, ne sont pas américanistes. Pendant la période prévue pour réaliser cette enquête, je n'ai réussi à m'entretenir qu'avec un seul homme (archéologue). Avec les autres hommes qui avaient été contactés, nous n'avons pas réussi à trouver de moment libre convenant aux deux parties.

Par ailleurs j'ai pu m'entretenir avec cinq secrétaires de rédaction (trois travaillent pour le groupe *Gallia*, une pour la revue *Paléorient*, la dernière pour la *Revue archéologique* et *Syria*) et trois directeurs de revue (*Gallia préhistoire*, *AES* et *Paléorient*). Notons que ces directeurs de revue sont aussi parfois des auteurs d'articles. Ils ont donc été questionnés aussi bien à propos de la revue pour laquelle ils travaillent, que sur les mots-clés qu'ils ont pu fournir au moment d'une publication.

Le questionnaire réservé aux secrétaires de rédaction et aux directeurs/rédacteurs de revue portait également sur le thème de la valorisation des revues. Il était précisé que, par valorisation, il fallait entendre tout ce qui permet de rendre plus visible et plus accessible le contenu d'une revue. Il s'agissait donc d'essayer de déterminer les besoins en matière de valorisation pour l'ensemble des revues du pôle. Cet aspect sera abordé dans la troisième partie de ce mémoire.

L'enquête composée de ces questionnaires et d'observations réalisées chaque jour dans le cadre de mon activité professionnelle a pu être menée dans le courant des mois d'avril à juin 2009.

6.2 Mots-clés et résumés : plusieurs points de vue

6.2.1 Du côté des auteurs

La première impression et/ou réaction à l'annonce de mon sujet d'enquête a été globalement : « ah ! oui, c'est intéressant ! ». La deuxième était : « oh ! mais tu sais, je n'ai pas grand-chose à dire sur les mots-clés ». Ouverture, puis fermeture, résultat peut-être d'une légère gêne ? Une gêne due au fait qu'ils n'accordent généralement que très peu de temps, dans leur réflexion, aux mots-clés. Une gêne non justifiée car, de leur côté, la plupart

des revues ne consacrent elles aussi que peu de temps aux mots-clés. Cette première observation est donc intéressante.

La plupart des auteurs rencontrés affirment qu'ils ne pensent aux mots-clés qu'après la rédaction de l'article, voire quand la rédaction de la revue le leur demande. Le résumé est d'ailleurs réalisé au même moment, à la fin, bien après la conception et l'écriture du texte.

Pour construire les mots-clés, les auteurs rencontrés se posent un certain nombre de questions : faut-il mettre les mots du titre dans les mots-clés ? doit-on donner des termes généraux, spécifiques, des expressions-clés ? De plus, le nombre de mots-clés demandés pose parfois un problème car, selon les auteurs interrogés, il faudrait pouvoir en donner plus que cinq ou six car ce n'est pas suffisant pour rendre bien compte du contenu d'un texte.

> J'ai l'impression qu'il y a des choses obligatoires, le lieu, la problématique générale. En fait, le plus gros problème c'est le nombre de mots-clés. Quand on te demande de donner seulement cinq mots-clés, tu mets forcément le lieu, la problématique générale, ensuite il reste trois places, mais ce n'est pas suffisant. Je pense que, pour mes thèmes de recherche, il ne faut pas utiliser des choses trop, trop précises. C'est difficile la balance entre le général et le particulier dans le choix, mais la contrainte pour moi c'est surtout la limitation du nombre de mots-clés et l'obligation de préciser le lieu. Quand je publie dans une revue d'ethnologie proprement dit, je ne mets pas ethnologie ; en revanche je mets parenté, migration. Pour le lieu, je mets le pays ou la région. J'ai travaillé en Nubie égyptienne. Donc si j'ai suffisamment de places, je mets Nubie et Égypte. Sinon, je mets juste Nubie et, à ce moment-là, cela pourrait devenir une discussion avec mon correspondant dans la revue qui pourrait me dire que personne ne connaît la Nubie ou que cela renvoie nécessairement à l'époque pharaonique. [...] Dans la revue Ethnologie française, j'ai participé au numéro sur les mémoires plurielles. Je n'ai pas utilisé le mot ethnologie, ni le mot mémoire dans mes mots-clés, ce dernier était en plus dans mon titre. On ne peut pas reprendre les mots du titre. En même temps, j'essaie de trouver des mots-clés et pas des expressions-clés, ce qui est encore plus compliqué parce que si je mets « immigré africain », est-ce que c'est vraiment précis ? Non ça a l'air précis, mais en réalité ça ne l'est pas du tout. Je travaille par exemple sur les problèmes de transmissions ou de non transmissions de génération en génération d'un dispositif de parenté chez des migrants, plus exactement chez des immigrés et des migrants qui ne sont pas encore stabilisés. Ce sont des africains, mais des africains de l'ouest et du centre. Donc ce n'est pas toute l'Afrique. Jusqu'où a-t-on besoin d'être précis ? Plus tu mets des mots-clés, plus cela peut devenir redondant et trop particulier, du coup cela ne sert plus à rien. Dans le numéro thématique d'Ethnologie française auquel j'ai participé, j'ai aussi regardé ce que les autres avaient mis comme titre, même si ce n'était que des titres provisoires, pour savoir si le mot transmission, puisque mémoire je ne pouvais pas l'utiliser, avait un sens. Après j'ai des choses comme liens de parenté et relations de parenté qui, pour moi, ne veulent pas dire la même chose et je dois sans arrêt m'en expliquer dans les textes puisque cela correspond à ma manière de voir. Est-ce que, dans les mots-clés, on doit se limiter à des choses générales qui vont donner une orientation générale et ensuite les gens vont aller voir ce que tu avances en particulier ou alors tu peux mettre quelque chose de très particulier qui renvoie vraiment à ta manière de penser et à ton analyse et, à

> ce moment-là, pour les gens cela ne dit rien du tout. Par exemple, sur le mariage arabe, je peux toujours le mettre. Mais moi ce que je prétends, c'est que, pour analyser le mariage arabe, le plus important c'est la relation frère-sœur. Donc il faut que je puisse mettre les deux et le lien entre les deux se trouve dans l'article. Je ne peux pas trouver un mot-clé ou une expression-clé qui rende compte de mon positionnement parce que cela n'existe pas. (Ethnologue, chargée de recherche au CNRS)

Ce long extrait d'entretien illustre parfaitement les questionnements posés et les problèmes rencontrés par les auteurs dans l'élaboration des mots-clés. Il va également plus loin en proposant une réflexion plus fine sur l'indexation proprement dite. Revenons sur certaines des remarques énoncées.

À propos des mots-clés géographiques, ces derniers pouvaient laisser espérer moins d'interrogation quant à leur constitution, puisqu'il s'agit de préciser un lieu tout simplement. Pourtant on constate, ici comme dans d'autres entretiens, que ce n'est pas si simple de savoir quel degré de précision il faut donner : faut-il mentionner le pays [24] ou la région ou la ville ou le village… ? Doit-on se limiter à une seule indication ou être plus précis ? D'après certains auteurs, cela dépend de l'orientation du texte et de la perspective mises en avant. Très bien, mais alors peut-on être sûr en donnant une précision géographique très fine que les gens qui vont faire des recherches vont connaître cette spécificité ? Pour les auteurs archéologues, la localisation spatiale est essentielle, voire primordiale ; elle va de pair avec la chronologie, la localisation temporelle. La localisation spatiale, dans le contexte archéologique, renvoie au nom des sites étudiés ou mentionnés dans le texte [25]. Quoi qu'il en soit, tous s'accordent pour dire qu'il faut au moins un mot-clé géographique pour situer l'aire étudiée. Le degré de précision dépend vraiment de l'axe choisi dans le texte, mais plus on voudra être précis et plus il sera nécessaire de mettre des mots-clés généraux permettant de situer le terme spécifique [26].

---

24. Signalons rapidement le problème des pays qui ont changé de nom au cours de leur histoire. C'est le cas par exemple de la République de Haute Volta, en Afrique, proclamée le 11 décembre 1958, qui est devenue le Burkina Faso le 4 août 1984, soit 24 ans après son indépendance.
25. Notons que, pour les sites archéologiques, un problème de confidentialité lié aux risques de pillage du site empêche les auteurs de donner les coordonnées exactes du lieu fouillé.
26. Dans le cas d'une indexation proprement dite, il pourrait être intéressant d'envisager un système permettant l'autopostage. Il s'agit d'un « procédé permettant d'effectuer automatiquement une indexation complémentaire d'un document ou d'une question par tous les descripteurs appartenant à la même branche de l'arborescence du thésaurus que le descripteur le plus spécifique utilisé lors de l'indexation. L'autopostage générique (vers un niveau supérieur) peut être effectué lors de l'indexation et lors de la recherche. L'autopostage spécifique (vers un niveau inférieur) s'effectue lors de la recherche » ([1] Boulogne).

Le contexte de publication peut conditionner la constitution des mots-clés. Il est clairement dit dans cet extrait que le mot « ethnologie » n'est pas mis dans les mots-clés quand la publication se fait dans une revue d'ethnologie proprement dite. Le mot « mémoire » n'est pas non plus mentionné quand l'article en question est prévu pour un numéro spécial dont le titre contient le mot « mémoires ». Ainsi la répétition des mots du titre (que ce soit le titre de la revue, du numéro spécial ou encore de l'article) dans les mots-clés pose problème à certains auteurs tandis que d'autres décident de ne pas être redondants.

On voit aussi, à travers cet extrait, les problèmes que posent les mots-clés thématiques. Faut-il rester dans les aspects généraux ou entrer dans ce qu'il y a de véritablement spécifiques ? comment rendre compte de notre pensée ? un mot ne suffit pas, une expression n'est pas toujours parlante. L'exemple du « mariage arabe » est assez explicite. Il serait inconcevable en ethnologie d'indexer un texte qui parle du mariage arabe en mettant comme mots-clés « mariage » d'un côté et « arabe » de l'autre. En effet, il s'agit d'un concept très précis en ethnologie, relevant des études sur la parenté [27]. Mais l'auteur s'interroge pour savoir s'il faut donner dans les mots-clés le fond de sa pensée, à savoir dans ce cas précis qu'il envisage le mariage arabe à travers la relation frère-sœur. Ou faut-il juste expliciter les thèmes abordés, ce qui ne remplace en rien la lecture de l'article, mais permet d'en connaître l'objet d'étude ?

Remarquons, par ailleurs, que certains auteurs interviewés ont tranché sur ces questions et s'imposent, en quelque sorte, des règles. Celles-ci ne sont validées que par l'auteur lui-même. Celui-ci s'empresse d'ailleurs de préciser que, comme aucune indication n'est donnée dans les rédactions de revue, ces règles permettent de se forger un cadre de réflexion pour les mots-clés.

> En général, dans mes mots-clés, j'en mets deux généraux et le reste se réfère à des notions plus spécialisées (le type d'artefact ou d'objet de réflexion sur lequel porte l'article, les notions discutées…). Je fais en sorte de mettre des mots différents de ceux du titre afin d'éviter toute redondance. Je mets des thèmes très spécifiques, les points accrocheurs de mon article, ce qui n'apparaît pas dans le titre et, pourtant, qui sont dans le texte. J'essaie d'être complémentaire au titre. Je mets des mots-clés thématiques, mais aussi la région et la période s'ils ne sont pas déjà précisés dans le titre. (Archéologue, chargée de recherche au CNRS)

27. « Le mariage arabe, au sen strict, désigne l'union des cousins parallèles patrilatéraux [mariage entre les enfants d'un couple et les enfants du frère du père] vrais ou classificatoires ; du point de vue de la typologie des structures d'alliance, il exprime la préférence pour le mariage au sein de la lignée agnatique [apparenté exclusivement par les hommes] » (Barry Laurent et al., « Glossaire de la parenté », L'Homme, 154-155, 2000, pp. 721-732).

Ainsi il est intéressant de constater qu'un certain nombre d'aspects doit être précisé : l'aire géographique, le groupe étudié, la chronologie pour les archéologues, les thèmes principaux et secondaires si on a assez de place. En donnant ces indications, on se met, en quelque sorte, dans la peau de l'éventuel lecteur du texte. C'est ce que dit un directeur de revue qui répond en tant qu'auteur d'un texte, et non pas en tant que directeur de revue :

> Pour les construire, je me pose la question suivante : qui peut avoir intérêt à chercher mon article, c'est quelqu'un qui s'intéresse à quelle question ? J'essaie de me mettre dans la tête de l'utilisateur et je n'essaie absolument pas de, et je suis contre vouloir, mettre en mots-clés tout ce qui peut incidemment passer dans l'article. Je ne veux pas perdre mon lecteur. Si, dans l'article, il y a par exemple un peu de descriptions ethnographiques sur des passages piétonniers, je ne vais pas pour autant mettre « passages piétonniers » en mots-clés. (Directeur de revue, Sociologue, retraité)

Dans ce cas précis, on voit que l'auteur n'a pas pour but que son texte soit lu par une majorité de lecteurs, mais il préfère viser des personnes susceptibles d'être intéressées par le véritable sujet de son texte. D'autres auteurs préfèrent voir plus grand et tentent donc de toucher un public plus large en proposant des mots-clés généraux.

Par ailleurs, certains auteurs prennent beaucoup plus de soin à trouver un titre qu'à élaborer des mots-clés parce que le titre d'un texte donne plus d'informations directes que des mots-clés, perçus comme restrictifs. De plus, on est sûr qu'un titre sera lu, on en est moins sûr pour les mots-clés. Un titre doit donc être « attirant », selon le mot d'un auteur.

Penchons-nous à présent sur la question : à quoi servent les mots-clés selon les auteurs ? Plusieurs n'en voient pas l'utilité, ils avouent même en négliger la fabrication ; certains ne savent pas à quoi ils peuvent bien servir. Un auteur parle de classement : « les mots-clés servent à classer aux bons endroits afin que le texte soit repérable facilement » (Ethnologue, chargée de recherche au CNRS). Un autre précise qu'ils peuvent être utile à l'indexation, une indexation servant à trouver un document dans une masse de documents : « ils doivent servir à indexer le texte pour que d'autres chercheurs ou personnes intéressées par le sujet le trouvent dans la masse d'articles existants » (Docteur en ethnologie). Enfin, un dernier évoque l'accès à l'information : « les mots-clés, c'est pour accéder aux textes, le résumé c'est ce qu'il y a dans le texte, il donne le contenu » (Ethnologue, chargée de recherche au CNRS). Les mots-clés sont donc des points d'entrée pour arriver jusqu'au texte, le résumé en revanche informe sur ce que contient ce même texte.

Le rapport résumé/mots-clés n'est pas toujours évident. Si certains auteurs voient dans les mots-clés un complément au résumé, d'autres ne les trouvent pas très utiles, ils ne donnent que très peu d'information, il est préférable de lire le résumé.

Pour la rédaction du résumé, un des auteurs rencontrés explique que :

> C'est bâclé à la fin. Le résumé n'a rien à voir avec la proposition d'article que j'ai faite au départ. Je ne peux m'en servir pour le résumé parce qu'un résumé c'est plus court, cela n'a pas la même forme. J'y consacre peu de temps et souvent dans l'urgence. En général, j'essaie de faire le résumé le plus simplement possible. J'ai toujours l'idée du suspens parce que j'aimerais bien attirer le lecteur par le résumé, mais je pense que je n'y arrive pas, donc c'est plat, ça dit ce qu'il y a dans l'article : le contexte et les propositions. J'essaie d'y mettre ce qui, pour moi, en tant que lecteur, pourrait m'attirer. (Ethnologue, chargée de recherche au CNRS)

L'idée de « suspens » évoqué dans ce passage est loin des résumés tels qu'on les conçoit en documentation. L'auteur aimerait composer son résumé comme un texte de quatrième couverture, il se place donc davantage dans un esprit proche de celui d'un éditeur commercial qui veut attirer des lecteurs pour vendre, bien que, dans le cas de l'auteur, il ne s'agit pas d'attirer pour vendre, mais pour être lu. Finalement, ce procédé très tentant n'est pas envisageable dans les faits, car il demande du temps. Il vaut mieux s'en tenir à un résumé classique.

Certains des auteurs rencontrés considèrent que les mots-clés et le résumé se complètent. D'autres, en revanche, ne voient pas l'intérêt des mots-clés, ils ne le lisent jamais ou très rapidement, juste parce qu'ils ont l'habitude de tout lire. Ils préfèrent regarder le titre et le résumé qui sont beaucoup plus informatifs.

6.2.2 Du côté des directeurs de revues

Dans les revues présentes à la Maison René-Ginouvès, le soin apporté aux mots-clés est très succinct car les mots-clés comme le résumé sont considérés comme la propriété de l'auteur.

> Nous évitons d'y toucher car ils font partie de la pensée des auteurs. De même que l'on ne réécrit pas un texte, on ne touche pas aux mots-clés des auteurs. C'est le respect scientifique du droit d'auteur. Les auteurs sont libres. La rédaction a juste un droit de regard et n'intervient dessus qu'en cas de gros problème. (Directeur de revue)

J'apporterai quelques nuances à cette affirmation. D'après mon expérience personnelle dans la revue pour laquelle je travaille, il est possible de proposer des modifications pour les mots-clés et le résumé. Il est également possible d'améliorer la rédaction du texte en réécrivant des passages difficiles à lire. Mais toutes ces modifications ne peuvent être envisagées, sans avoir été au préalable soumises à l'auteur qui devra donner son accord pour introduire ces ajustements. Aussi l'intervention du secrétariat de rédaction sur l'ensemble de la contribution de l'auteur dépend de la politique choisie par le comité de rédaction et/ou le conseil scientifique. Dans une autre revue, le directeur explique qu'il

n'hésite pas à éliminer des termes si ceux-ci lui paraissent inutiles ou si l'auteur en a mis un trop grand nombre.

Pour un des directeurs de revue, le résumé doit être vraiment peaufiné car, si une personne peut lire quelque chose dans un article, c'est bien le résumé. Il est donc nécessaire d'y trouver, certes en version réduite, tout ce qui a été abordé dans le texte. Ce directeur convient alors que les résumés peuvent être longs, ce qui n'est pas gênant dans la mesure où toutes les idées développées dans le texte y sont bien présentes. Pour les mots-clés, en revanche, il leur accorde moins d'attention.

Pour un autre directeur, les mots-clés et le résumé forment un tout. Le résumé donne l'articulation des idées dans le texte ; les mots-clés la quintessence d'un article. Une secrétaire de rédaction compare les mots-clés à un « squelette » ou à une « radiographie » de l'article. Il est important pour ces personnes de lire non seulement le titre et le résumé, mais aussi les mots-clés.

> Le résumé ne suffit pas à rendre compte de l'article, mais avec les mots-clés on peut avoir une bonne idée. Ils traduisent les aspects de la pensée d'un auteur. Ils sont très hétérogènes car liés à l'auteur. Ils peuvent aider à l'indexation qui se fait grâce à une liste de mots limités, donc si dans les mots-clés on en voit des pertinents, on peut s'en servir pour indexer les articles dans la base des articles publiés. (Directeur de revue)

On voit dans ce passage que les mots-clés sont importants parce qu'ils donnent une bonne idée du contenu d'un texte et aussi parce qu'ils aident à l'indexation. Il existe déjà une liste limitée de mots permettant d'indexer les articles de la revue. Cette liste peut être enrichie par les mots donnés par l'auteur du texte.

6.2.3 Du côté des secrétaires de rédaction

D'une façon générale, les secrétaires de rédaction (ou SR) interviennent très peu sur les mots-clés. Elles corrigent les éventuelles coquilles, proposent d'en supprimer quand il y en a trop ou d'en ajouter quand il manque certaines informations. Pour certaines secrétaires de rédaction, les mots-clés doivent nécessairement couvrir les thèmes principaux, l'aire géographique, la période, le nom du site archéologique ou le groupe étudié.

La plupart des SR perçoivent un problème dans les mots-clés parce qu'ils sont trop libres, autrement dit aucune règle n'est fournie aux auteurs pour les constituer. Ce problème ne permet pas de les harmoniser : aucune concordance entre les différentes orthographes d'un même mot, pas de précisions sur l'emploi du pluriel et/ou du singulier, pas d'indications sur les types de mots-clés importants pour un article en ethnologie ou en archéologie.

Selon une des secrétaires de rédaction, les mots-clés sont très importants surtout pour les étudiants car les chercheurs confirmés ont plutôt tendance à chercher par titre ou auteur alors qu'un étudiant qui ne sait pas nécessairement ce sur quoi travaille tel ou tel chercheur a besoin des mots-clés pour trouver des textes traitant d'un sujet proche du sien. C'est aussi très utile pour les post-doctorants qui commencent à publier des textes et dont le travail n'est pas encore connu dans la communauté scientifique. Dans ce cas précis, les mots-clés pourront aider à faire connaître leurs écrits, en permettant de les trouver facilement.

Selon d'autres SR, les mots-clés servent à donner une vue synthétique de l'article, mais aussi à faire connaître des textes et à les ranger dans de grandes catégories pour les retrouver plus facilement. Il est nécessaire d'exercer un contrôle sur ces mots-clés soit en utilisant des langages déjà existants, soit en créant un vocabulaire propre à une ou plusieurs revues. Je reviendrai sur ces différentes idées dans la troisième partie.

7 Recherche d'informations

Je vais, à présent, aborder la recherche d'informations aussi bien dans le domaine de l'archéologie et de l'ethnologie en général que dans une revue spécialisée dans un de ces domaines. Une des parties de l'enquête concernait justement les méthodes mises en place par les chercheurs pour chercher et trouver des informations dans leur domaine. Après avoir fait le point sur les réponses obtenues, je présenterai plus précisément les outils papier et électronique qui existent pour faire une recherche d'informations dans une revue spécialisée.

7.1 Typologie, outils et méthodes

Grâce à l'enquête menée auprès de quelques chercheurs, je vais tenter de répondre aux questions suivantes : quel type d'information, les chercheurs sont-ils conduits à rechercher ? comment font-ils une recherche d'informations en général ? Comment s'y prennent-ils pour trouver des informations dans une revue spécialisée ?

7.1.1 Typologie des informations recherchées

Les informations recherchées sont de plusieurs types : des données, des analyses et/ou interprétations, des références bibliographiques, des textes (articles, livres…), des illustrations (images, photos, dessins…).

> Cela dépend de l'objectif :
> -des données précises sur une région ou un thème, contenues dans les articles ;
> -la méthodologie employée par les auteurs ;
> -les types d'interprétations, l'articulation des données pour émettre un interprétation ;
> -la bibliographie pour la compléter et être sûr de ne pas passer à côté de choses importantes ;
> -des illustrations (un plan de site [archéologique], une image…).
> (Archéologue, chargé de recherche au CNRS)

La recherche de références bibliographiques et l'accès aux textes sont un préalable nécessaire à une recherche d'informations sur un sujet, sur un point précis, sur la pertinence d'un concept. Ces informations vont permettre ensuite d'accéder à des données, des analyses et/ou interprétations, parfois même à des illustrations.

> Je peux vouloir compléter une recherche bibliographique sur un point précis ou alors chercher à me faire une idée sur un sujet donné, dans le cadre d'une recherche ou d'un cours. (Docteur en ethnologie)

Ces recherches d'informations se font généralement pour écrire un article, un ouvrage ou pour construire un cours, une conférence, un séminaire.

> Le plus souvent, c'est pour la rédaction d'un article ou d'une publication. Parfois cela peut aussi être pour rédiger un projet, mais c'est moins fréquent : cela permet de se positionner sur le thème en fonction de ce qui a déjà été fait.
> Il y a des thèmes qui m'intéressent, sur lesquels j'ai déjà travaillé à un moment donné et je continue à rassembler des informations dessus. Par exemple tout ce qui est en relation avec l'archéologie funéraire, le traitement des squelettes, des restes humains. Dans le cadre de ma thèse, j'avais travaillé sur le thème des propulseurs, donc quand je tombe sur un article sur les propulseurs, je le mets de côté, aujourd'hui encore, même si je ne le fais pas de manière systématique.
> D'autres fois, je collecte des informations pour une recherche en cours (pour augmenter ma bibliographie) et/ou sur une thématique qui m'intéresse.
> Je le fais aussi dans le cas d'un cours à faire où on présente des données qu'on connaît relativement bien, mais que l'on doit replacer dans un contexte plus large. Sur des thèmes pour lesquels on est moins à l'aise ou pas spécialement familiarisé, cela implique d'aller chercher des références moins générales, plus spécifiques.
> À propos d'un cours sur l'idéologie funéraire, chacun devait traiter d'un thème. Moi j'ai choisi le thème même de l'idéologie funéraire même si je savais qu'il était très large, je l'ai fait aussi pour en apprendre un peu plus. J'avais un certain nombre d'exemples provenant de mes propres travaux ou d'autres travaux qui avaient été fait sur la Mésoamérique. Un exemple, cela correspond à une étude de cas. J'ai dû aussi aller chercher des bouquins généraux sur le thème, sur les pratiques funéraires pour trouver des informations sur le lien entre pratiques funéraires et idéologie, voir comment ces deux aspects avaient été perçus par certains auteurs. Et puis, comme il s'agissait d'un cours d'archéologie, on essaie toujours de faire des trucs visuels, donc j'ai essayé d'aller pêcher des images pour illustrer certains propos ou des exemples que j'illustrais par une image. (Archéologue, chargé de recherche au CNRS)

À travers ce témoignage, le type d'informations recherchées et le cadre dans lequel elles peuvent être recherchées sont clairement précisés.

Un des chercheurs rencontrés explique qu'il tente souvent de trouver des informations récentes et spécialisées pour savoir ce qui se fait actuellement, pour connaître les dernières découvertes, bref pour se tenir informé de l'actualité de sa discipline.

De cette rapide typologie des informations recherchées découle nécessairement un questionnement sur les méthodes de recherche d'informations. On constate grâce à ces témoignages qu'il existe également une typologie de la recherche d'informations.

-En tant qu'étudiants et/ou chercheurs, on peut en effet être amené à faire une **recherche de type ponctuel** [28]. Elle répond à un besoin limité dans le temps. C'est le cas par exemple de la préparation d'un cours, d'une conférence, d'un exposé…

-Il est aussi possible de mettre en place une **recherche de type continu**. Elle permet de répondre à un besoin d'être tenu régulièrement informé sur un sujet précis, souvent annexe au thème d'étude. C'est ce qui se fait, par exemple, pour connaître l'actualité d'une discipline ou d'un domaine, une découverte…

-Enfin on peut aussi considérer une **recherche de type courant**. Elle répond à un besoin de suivi des parutions et/ou manifestations concernant son propre domaine d'étude, sa spécialisation. C'est ce qui se pratique donc au quotidien par un chercheur sur sa spécialité.

Voyons à présent les outils et méthodes utilisées pour trouver des informations.

7.1.2 Faire une recherche d'informations en général

Selon la plupart des chercheurs interrogés au cours de cette enquête, il semble que le premier réflexe pour chercher de l'information passe par Google, Google Scholar et Google image.

> Je fais mes recherches de deux façons qui se complètent. Je cherche sur Google à partir de mots-clés. J'obtiens des réponses (textes, articles ou livres) qui me permettent d'approfondir et de préciser ma recherche. Donc je pars de ce que je trouve pour affiner cette recherche. Je regarde ensuite les bibliographies de ces documents et je fais d'autres recherches en fonctions de ça. Je me suis fait une alerte sur Google à partir de deux mots-clés : décohabitation et polygamie. Je cherchais des cas car j'en avais sur mon terrain. Grâce à cette alerte, j'ai régulièrement des infos, mais certaines répètent des choses que je connais depuis un moment. Mais j'ai quand même eu accès à un mémoire d'une assistante sociale sur ce sujet. (Ethnologue, chargée de recherche au CNRS)

Dans cet extrait, il est fait mention des mots-clés. Une des personnes interrogées a clairement précisé que les mots-clés donnés pour un article qu'elle avait écrit n'avait rien à voir avec ceux qui lui servaient à la recherche d'informations. Les premiers décrivent le contenu du texte, les seconds aident à la recherche d'informations. Il est intéressant de noter cette démarcation entre les mots-clés qui décrivent le contenu et ceux utiles à la recherche.

---

28. Les types définis ici ont pu être ainsi qualifiés grâce à mes observations de terrain et aux commentaires recueillis pendant les entretiens ou au cours de discussions informelles.

Un autre chercheur explique sa méthode de recherche en ces termes :

> Pour faire une recherche sur les pratiques funéraires, je suis d'abord allé consulter les bouquins de référence ou les articles qui sont cités par tous et tout le temps. Je vais voir ces articles et ces bouquins pour voir qu'elles sont les différentes théories ou explications ou approches qui sont proposées. Puis je vais chercher des exemples concrets pour expliciter, pour illustrer chacun de ces points de vues. Je cherche d'ailleurs des images pour illustrer ces exemples. Je dépouille aussi la bibliographie d'un ouvrage et d'un article pour consulter les références signalées si elles me sont utiles. Pour un thème que je ne connais pas du tout, je vais voir un collègue spécialisé dans le domaine, je ne pars pas sans rien, face à l'étendue des écrits sur le sujet. (Archéologue, chargé de recherche au CNRS)

Cette méthode est assez traditionnelle : prendre les ouvrages de référence sur un sujet précis et dépouiller la bibliographie, puis à partir de cette bibliographie, consulter des textes mentionnés, regarder leur bibliographie, et ainsi de suite…

Un autre chercheur a mis en place quelque chose qui pourrait bien ressembler à un système de veille :

> Je fais une veille sur toutes les revues qui m'intéressent. Je regarde leur sommaire sur Internet, sur des sites de revues, sur biblioSHS ou encore à la bibliothèque d'archéologie. Je me fais aussi des liens RSS pour les revues. En fait ce sont des liens sur leur sommaire. Je me suis inscrite sur des sites d'éditeurs pour les revues et les livres. J'ai choisi les thèmes qui m'intéressent comme archéologie, art, cognition, histoire des sciences, éthologie, géographie, antiquité, ethnologie, anthropologie ; je cherche des éditeurs spécialisés sur des régions spécifiques (pour moi, c'est l'ex-Union soviétique, Asie, Proche et Moyen Orient). Les revues que je consulte sont souvent mensuelles ou bi mensuelles. Je ne le fais pas pour les revues françaises car, là en général, je suis informée ici par des collègues. Je le fais surtout pour des revues étrangères, notamment des revues anglo-saxonnes. V. nous envoie la liste détaillée (avec sommaire) des dernières acquisitions de la Bibliothèque d'archéologie concernant l'Orient ancien (qui comprend l'Asie centrale). Et c'est très utile. Dans un article que je suis en train de lire, je regarde aussi la bibliographie même si c'est des références qui ne sont pas toujours récentes. Puis, parfois, les collègues m'envoient leur article en pdf. (Archéologue, chargé de recherche au CNRS)

On voit ici un usage systématisé des sites de revues, des catalogues électroniques de bibliothèques, des sites d'éditeurs. L'importance des collègues comme source d'informations est, comme dans l'extrait précédent, mentionnée.

Dans d'autres extraits d'entretien, il apparaît que, même si, pour certains, le premier réflexe est d'aller sur Google, l'étape suivante de la recherche se fait en utilisant des outils

spécialisés comme JSTOR, biblioSHS, Science direct, Revues.org, Cairn, Persée, catalogues de bibliothèques spécialisées... [29]

Ce qui précède a permis de montrer quelques usages en matière de recherches d'informations dans le domaine des sciences humaines et sociales, plus précisément en ethnologie et archéologie.

Il peut être intéressant ici de distinguer les usages, ce que l'on a coutume de faire, et les besoins, ce qui serait nécessaire d'avoir. En effet, en regardant ce que l'on a l'habitude de faire pour chercher des informations, il sera alors possible de suggérer ce qui serait nécessaire d'avoir pour améliorer cette recherche [30]. Des propos recueillis, il est souvent ressorti que le temps consacré à la recherche d'informations ou à se documenter sur le domaine d'étude ou sur un domaine annexe est assez limité. Pourtant, le besoin d'avoir l'information et d'y accéder de façon rapide et aisée se fait sentir.

L'impression générale qui ressort de l'ensemble de ces propos est qu'il semble, en fait, davantage nécessaire pour les chercheurs d'être informés, plutôt que de devoir ou de savoir chercher l'information. Améliorer la diffusion, le signalement et l'accessibilité de l'information pourrait donc répondre à un besoin latent des chercheurs. On peut ici introduire la notion de service car, en proposant de telles améliorations, il serait intéressant de réfléchir en termes de mise en place de services.

Avant d'aller plus loin, regardons la recherche d'informations dans une revue spécialisée.

7.1.3 Trouver une information dans une revue spécialisée

J'ai tenté de montrer plus avant qu'une revue scientifique pouvait être considérée comme un système d'informations organisées. Elle est spécialisée dans un domaine particulier, dans une thématique spécifique... Cela signifie qu'avant même de consulter une revue, un chercheur connaît préalablement le type d'information que l'on peut y trouver et le champ de recouvrement thématique qui en fait sa spécificité. Néanmoins, ce n'est pas parce que l'information est organisée et plus ou moins bien repérable que cela permet de répondre à tous les types de recherche.

---

29. Aucune des personnes rencontrées n'a mentionné les archives ouvertes du CNRS, soit HAL-SHS. Il semblerait que ces personnes ne déposent pas leurs travaux, ni ne consultent ceux déposés par leurs confrères.
30. On peut ici se reporter aux textes de Boyer ([14] Boyer) et Kolmayer ([15] Kolmayer), chacun abordant les usages et/ou les besoins des utilisateurs.

Les chercheurs interrogés ont une attitude variable face aux informations présentes dans une revue. S'il s'agit d'une revue reconnue dans leur domaine, ils la consultent très régulièrement [31].

> Soit c'est une revue que je consulte régulièrement et je me tiens au courant en la consultant parce que j'y suis abonné ou je vais la consulter en bibliothèque ou encore par Internet par les bases de données de revues en ligne comme BiblioSHS, Jstor ; soit un collègue m'en parle sachant que je n'ai pas le temps de consulter toutes les revues de mon domaine. (Archéologue, chargé de recherche au CNRS)

Les chercheurs questionnés disent regarder le sommaire dès sa parution, lire éventuellement les résumés des articles repérés dans le sommaire, voire même le texte si le résumé est vraiment pertinent par rapport à leur domaine de recherche.

> Sur papier, je feuillette la revue en regardant d'abord le titre, le résumé s'il y en a un, puis l'introduction. Puis si je suis intéressée, je regarde la structuration de l'article, les titres, et finalement la conclusion. En feuilletant, je vois les illustrations en passant et si elles me permettent d'en apprendre plus, tant mieux. Et finalement si j'ai en tête des références qui me semblent importantes, je regarde si elles sont citées en bibliographie.
> Sur le document électronique : je fais la même chose, souvent le résumé permet effectivement de se faire une idée, je regarde brièvement les mots-clés, mais pas vraiment, presque sans les voir je crois. En tout cas ce n'est pas une étape que j'identifie en soi. Si je recherche un mot plus précis, je fais une recherche par mot. (Docteur en ethnologie)

Finalement papier ou électronique, la procédure semble être la même. La seule chose qui fait la différence, c'est la possibilité de chercher à partir d'un mot précis dans la version électronique.

Les chercheurs peuvent aussi décider de regarder une revue pour rechercher des informations très spécifiques comme ce que tel auteur a écrit, ou ce qui a été publié sur tel sujet, tel concept, tel site archéologique…

> Quand je cherche des infos sur un sujet, je regarde d'abord qui a travaillé sur le même sujet. Puis je cherche dans les revues car elles donnent des infos récentes et très spécifiques, par exemple sur les dernières découvertes dans tel site. (Archéologue, chargée de recherche au CNRS)

On peut supposer qu'un chercheur confirmé a constitué, au fil de sa carrière, une grille de lecture qui permet d'évaluer l'information trouvée et un réseau de collègues lui fournissant

---

31. Dans cet esprit et pour répondre à ce type de besoins, la bibliothèque d'ethnologie et de sociologie comparative de Nanterre a mis en place une diffusion électronique des sommaires de revue, ce qui permet à l'utilisateur de prendre connaissance, de son bureau ou même de chez lui, du sommaire des derniers numéros parus et d'aller consulter, si la revue a une édition électronique, le résumé et/ou le texte quand celui-ci est en ligne.

les informations nécessaires à sa recherche, une sorte de « collège invisible » (invisible car fluctuant selon les thèmes). Un étudiant, même avancé, n'a pas toujours ce cadre de référents. Il doit donc mener des recherches « à tâtons », en utilisant au mieux les outils que les bibliothèques lui proposent, au pire des moteurs de recherche généralistes.

Quoi qu'il en soit, dans les deux cas, il semble probable que l'on puisse passer à côté d'une information pertinente, car ces méthodes de recherche empirique restent aléatoires et dépendantes de conditions précaires et/ou de personnes bien disposées.

Néanmoins, on peut dire que la recherche d'informations dans une revue spécialisée apparaît comme étant plus ciblée, plus spécifique. Quand on regarde dans une revue, c'est parce que l'on sait déjà plus ou moins quel type d'information on veut ou on peut trouver. Avec le développement de la mise en ligne des revues, on peut espérer avoir des outils plus performants non seulement pour faire des recherches, mais aussi pour être informé des nouvelles parutions, donc pour que l'information soit mieux diffusée et signalée.

Par ailleurs, aucune des personnes rencontrées n'a pensé à parler des index et/ou tables des revues. Si elles en parlaient, c'était parce qu'une question le suggérait et, pour la plupart, ces outils, tables et index, étaient très utiles pour retrouver un texte dans une revue avant le passage en ligne des revues. Or, depuis que ces dernières se sont mises au numérique, ces outils sont de moins en moins consultés. Donnons quand même quelques précisions à leur sujet et sur les résultats de recherche obtenus grâce à leur emploi.

7.2 Index, indexation et tables

Dans les revues scientifiques, des outils papiers étaient – pour certaines ils le sont encore – produits pour aider à la recherche d'informations sur l'ensemble des volumes publiés dans la revue en question. Ces outils papiers ont tendance à disparaître – même si, dans quelques revues, ce n'est pas le cas [32] – et sont remplacés par des outils électroniques.

Il s'agit principalement des index et des tables.

Les index de revue, en général, peuvent proposer plusieurs entrées : nom d'auteur, rubrique, thème, aire géographique, groupe ethnique… Les textes sont alors mentionnés avec le nom de l'auteur, le titre, le numéro de revue, l'année de publication et la pagination. Ces index se retrouvent parfois dans les versions électroniques, comme nous allons le voir pour le JSA.

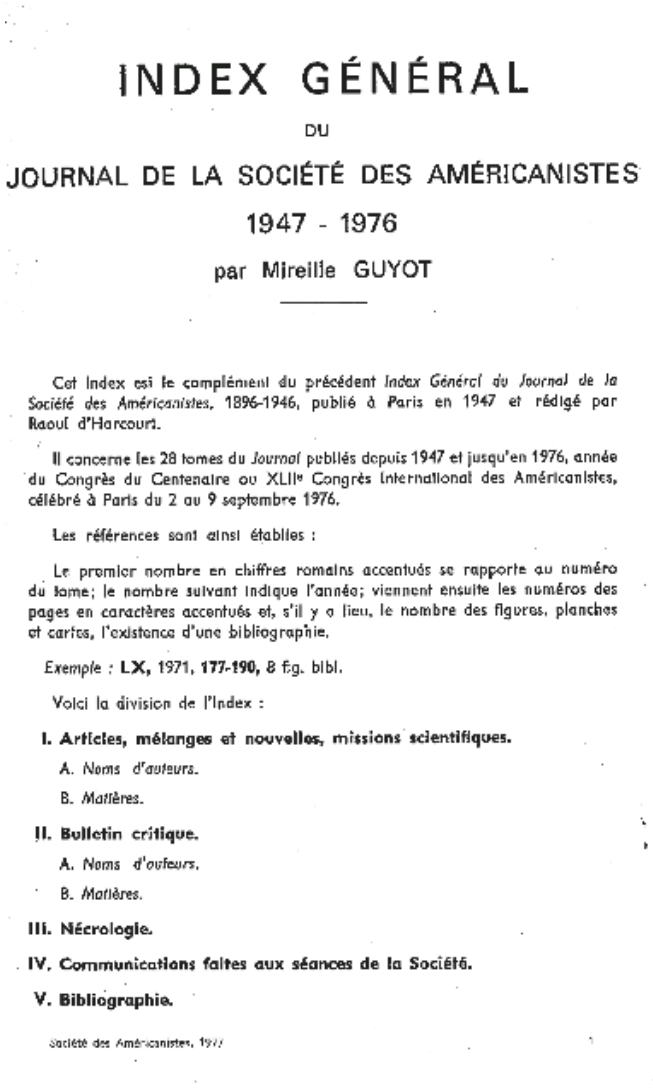
32. Ce n'est pas systématique car certaines revues tiennent à ces outils papiers. Il ne s'agit que d'une tendance qui va de pair avec la mise en place des versions électroniques des revues.

7.2.1 Le JSA

Pour le JSA, par exemple, trois index papiers ont été produits. Ils couvrent les périodes suivantes : 1896-1946, 1947-1976, 1977-1986. Ils se présentent tous à peu près de la même manière (voir page suivante). Dans le dernier index paru, celui de 1977-1986, quelques précisions sont formulées quant aux choix thématiques :

> L'index thématique de la partie correspondante aux articles est établi par discipline, géographique et par ethnie, langue ou culture. L'index par discipline réunit les textes en 7 groupes : Anthropologie biologique, Ethnohistoire, Ethnolinguistique, Ethnologie, Ethnomusicologie, Histoire et Préhistoire-Archéologie. A l'intérieur de chaque discipline on a déterminé, quand le nombre d'articles le rend possible, une division géographique. L'index géographique peut faire référence à des aires géographiques, à des pays ou à des régions. Par exemple : Andes, Mexique, Amazonie. L'index par ethnie ou par langue sert surtout à classer les travaux faisant partie des disciplines ethnologiques tandis que la classification par culture s'applique particulièrement aux articles d'archéologie. Dans les sections : Bulletin Critique et Mélanges et Nouvelles, l'index thématique est exclusivement géographique. Quant à la chronique d'information sur les Amérindiens, elle est présentée par pays. (*Journal de la Société des Américanistes*, Index général 1977-1986, pp. 327-328)

Ces précisions sont intéressantes. Néanmoins, il n'est fait à aucun moment mention des mots-clés, mais ce n'est pas parce que rien n'est dit qu'ils n'ont pas été utilisés à un moment donné, au moins comme source d'inspiration pour constituer cet index.



INDEX GÉNÉRAL

DU

JOURNAL DE LA SOCIÉTÉ DES AMÉRICANISTES

1947 - 1976

par Mireille GUYOT

Cet Index est le complément du précédent *Index Général du Journal de la Société des Américanistes*, 1896-1946, publié à Paris en 1947 et rédigé par Raoul d'Harcourt.

Il concerne les 28 tomes du *Journal* publiés depuis 1947 et jusqu'en 1976, année du Congrès du Centenaire ou XLIIe Congrès International des Américanistes, célébré à Paris du 2 au 9 septembre 1976.

Les références sont ainsi établies :

Le premier nombre en chiffres romains accentués se rapporte au numéro du tome; le nombre suivant indique l'année; viennent ensuite les numéros des pages en caractères accentués et, s'il y a lieu, le nombre des figures, planches et cartes, l'existence d'une bibliographie.

Exemple : **LX,** 1971, **177-190,** 8 fig. bibl.

Voici la division de l'Index :

I. Articles, mélanges et nouvelles, missions scientifiques.

A. *Noms d'auteurs.*

B. *Matières.*

II. Bulletin critique.

A. *Noms d'auteurs.*

B. *Matières.*

III. Nécrologie.

IV. Communications faites aux séances de la Société.

V. Bibliographie.

Société des Américanistes, 1977

Fig. 8 – Scan de la page d'introduction de l'index 1947-1976 du JSA.

Nous avons déjà vu que le JSA a une version électronique sur deux plateformes : Persée et Revues.org. Sur la première sont présentés les textes publiés dans les volumes allant de 1895 à 2000 : ils ont été scannés, puis ont fait l'objet d'une reconnaissance optique de caractères. Sur Revues.org se trouvent les textes publiés dans les volumes parus entre 2001 et aujourd'hui. J'ai déjà précisé comment se faisait la préparation pour cette mise en ligne dans les pages précédentes.

Sur la plateforme Revues.org (http://jsa.revues.org/), le JSA a une partie du menu vertical gauche qui est réservée à des index : auteurs, auteurs d'ouvrage, mots-clés, géographique/ethnique, thématique/disciplinaire, chronique. Il apparaît que l'indexation est mélangée avec ce qu'on pourrait appeler une catégorisation de l'information. En effet, les mots-clés, géographique/ethnique, thématique/disciplinaire correspondent à une sorte d'indexation. En revanche, auteurs, auteurs d'ouvrage, chronique renvoient davantage à des

types de textes (article, compte rendu, chronique), donc à ce que j'ai choisi d'appeler des catégories d'informations.



Fig. 9 – Copie d'écran des index du JSA sur Revues.org.

L'index appelé mots-clés dans la version électronique du JSA a été créé à partir des mots-clés d'auteurs et ne subit aucun contrôle. L'index géographique/thématique est réalisé à partir de ces mêmes mots-clés correspondant à ces thèmes.

Proposons quelques exemples de recherche et leur résultat :

Faisons une recherche rétrospective sur un texte, écrit il y a plusieurs années par Claude Lévi-Strauss, sur les Indiens bororo d'Amazonie. Utilisons les termes de recherche : « Lévi-Strauss » et « Bororo ». On peut choisir d'interroger en mettant le nom de l'auteur – mais sans pouvoir malheureusement spécifier qu'il s'agit du nom de l'auteur – et le nom du groupe ethnique – mais, là encore, sans pouvoir préciser qu'il s'agit d'un groupe ethnique. On obtient six réponses, sachant que celle qui nous intéresse vraiment n'apparaît qu'à la quatrième place.

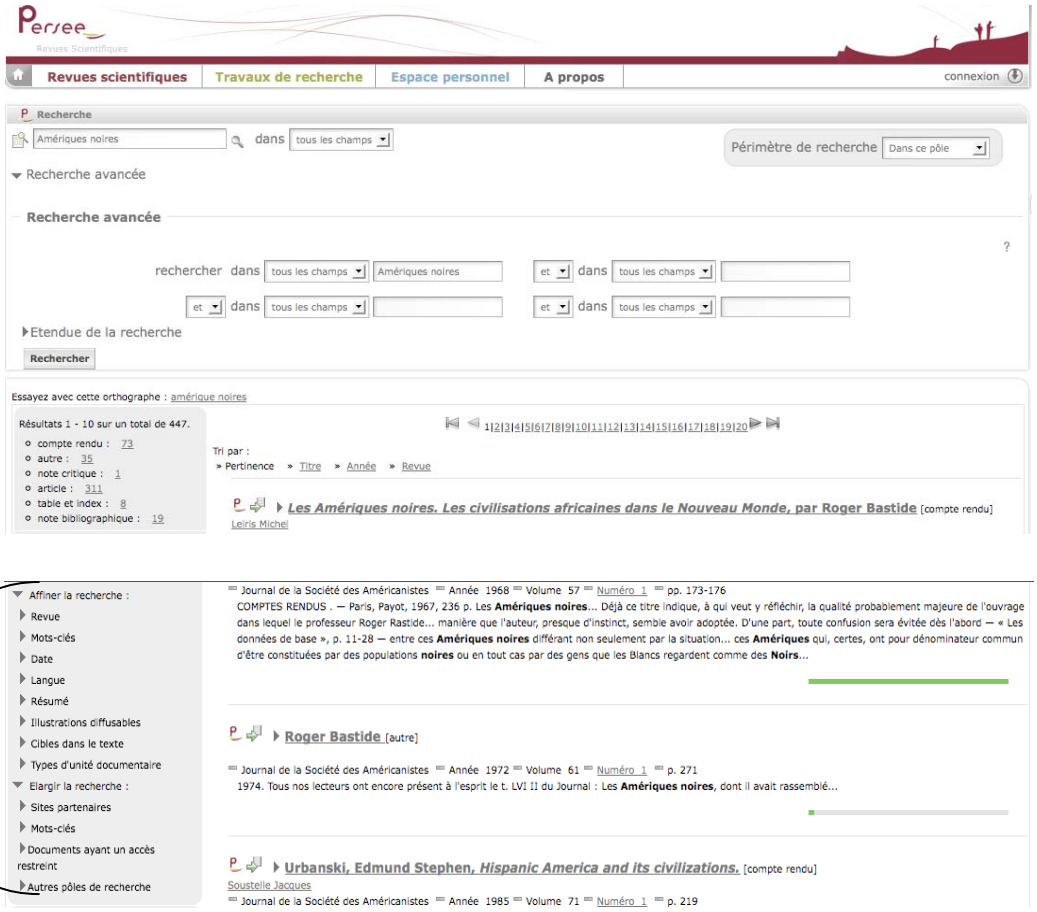


Fig. 10 – Copies d'écran de recherche sur Revues.org pour le JSA.

Le nombre de réponse diminue (trois seulement) si on interroge avec le nom et la date de publication, dans ce cas il faut donc savoir que ce texte a été publié en 1936.

Essayons à présent une recherche thématique. Cherchons ce qui a été écrit sur les Amériques noires dans le JSA depuis sa création. Utilisons l'expression au pluriel de recherche : « Amériques noires ».

Sur Revues.org, on obtient 17 réponses : elle n'inclut apparemment pas les textes qui sont sur Persée. Quand on fait la même recherche sur Persée, si on se met en recherche avancée et que l'on demande de ne chercher que dans les titres, on obtient une réponse. On fait la même chose, mais en demandant de ne chercher que dans le texte, on obtient 443 réponses. Enfin, en cherchant sur tous les champs, on trouve 447 réponses dont 311 articles.



Cette partie permet de préciser sur quelle revue doit porter la recherche, sur quels mots-clés, à quelle date, en quelle langue, sur les textes avec résumé, avec illustration…

Fig. 11 – Copies d'écran de recherche sur Persée pour le JSA.

Le nombre de réponses est très élevé et leur pertinence n'est pas toujours certaine. La recherche se fait sur l'intégralité des textes. De plus, l'absence de relations avec Revues.org est un dommage très gênant pour une personne souhaitant faire une recherche sur l'ensemble des numéros publiés. Pourtant le problème ne s'était pas posé pour la première recherche : serait-ce parce que les réponses étaient moins nombreuses ?

Quoi qu'il en soit, on constate que la recherche par le moteur de recherche proposé par Revues.org et/ou Persée n'est pas très satisfaisante. Elle présuppose que l'on sache déjà ce qu'il y a dans la revue, ce qui peut être vrai pour un chercheur, mais moins pour un étudiant. Ces moteurs ne permettent pas d'affiner véritablement les recherches. De plus, il est certain que le manque de contrôle sur les mots-clés et l'absence même d'indexation des textes sur l'ensemble de la revue ne facilitent pas les recherches.

7.2.2 La revue *Paléorient*

La revue *Paléorient* propose, sur son site web, une base de données des articles. Cette base de données offre la possibilité de faire des recherches par auteur, par titre, par mots du résumé, par mots-clés, langue, numéro de revue, année. L'interrogation peut se faire en français ou en anglais.



Fig. 12 – Copies d'écran de recherche sur le site de la revue Paléorient.

Les mots-clés proviennent d'une liste préalablement définie et pouvant être enrichie par les mots-clés donnés par les auteurs. Cette liste n'est pas à proprement parler un vocabulaire contrôlé. Certains mots-clés sont écrits de plusieurs façons différentes : on trouve par exemple des textes indexés avec carbone 14, d'autres avec 14C, d'autres encore avec Radiocarbone. En dépit de cela, cette base reste un outil très utile pour trouver des textes et des informations pertinentes. Pour la recherche d'informations dans cette revue, elle est un complément indispensable à la version électronique qui est sur Persée.

7.2.3 Des tables de revue

D'autres revues font des tables dont la périodicité peut être annuelle, biannuelle, quinquennale… Elle varie en fonction du rythme de parution de la revue. Ces tables sont publiées dans un des numéros de l'année et présentent tous les textes par ordre alphabétique du nom des auteurs, avec leur localisation dans la revue (numéro, pages) et, parfois, sous la rubrique dans laquelle ils sont parus.

Par exemple, la *Revue archéologique* publie, dans le second volume de l'année, une table reprenant les deux volumes. Elle est présentée par rubrique. Dans chacune des rubriques, sont mentionnés les textes rangés par ordre alphabétique du nom d'auteur, accompagnés de leur pagination seulement, sans le numéro de volume car cette revue a une pagination continue sur les deux volumes.

Ces outils ont (ou ont eu) le mérite d'exister et sont utiles pour retrouver un texte dans un volume, pour regarder ce qui a été publié au cours de l'année ou des dernières années, pour voir ce qu'un auteur a publié ces derniers temps… Ils permettent, certes, une certaine forme de mise en valeur des textes parus dans une revue, mais limitent, sans aucun doute, le type de recherches possibles : par exemple il n'est pas envisageable de faire des croisements thématiques pour trouver une information précise. La version électronique des revues aurait dû ou devrait remédier à ce manque…

*
* *

En prenant comme cadre de réflexion des revues spécialisées en SHS, dans leur version papier et électronique, j'ai donc essayé de montrer en quoi elles pouvaient être perçues comme des systèmes d'informations organisées, comment elles orientaient en raison de leur spécialité les recherches effectuées et comment les outils papiers et électronique qu'elles proposent étaient conçus et, par conséquent, pouvaient manquer d'envergure.

En choisissant d'entrer dans ces revues par la petite porte des mots-clés, nous avons pu décrire quelques pratiques des chercheurs et/ou étudiants pour l'élaboration de ces mots-clés et pour la recherche d'informations.

Il nous faut aller plus loin à présent en regardant ce que nous pouvons tirer de ces données recueillies.

Troisième partie : Valorisation des informations, valorisation des revues. Quelques pistes de réflexion

Dans cette partie, je vais tenter de tirer les fils des données exposées dans la partie précédente. En exposant les résultats de l'enquête menée au printemps 2009, j'ai tenté de décrire quelques pratiques et usages non seulement au niveau de la conception et de la représentation des mots-clés et résumés, mais aussi au niveau de la recherche d'informations.

J'ai également introduit rapidement une distinction entre usage et besoin. Cette distinction va, à présent, orienter cette troisième partie car, en adoptant le point de vues des besoins cette fois, après avoir regardé celui des usages dans la partie deux, il sera possible de réfléchir à d'éventuelles évolutions en vue d'améliorer l'indexation des textes publiés dans une revue, la recherche d'informations et, surtout, la visibilité et l'accessibilité des revues et des informations contenues dans celles-ci.

J'ai choisi de partager cette partie en deux sections : la première concerne l'indexation et les mots-clés. Elle commencera par un bilan des besoins rencontrés et/ou suggérés au cours des entretiens, également grâce à des observations faites pendant mon travail quotidien. De ce tableau, il pourra être intéressant de faire ressortir des propositions d'amélioration dans la constitution des mots-clés. Elle contiendra une présentation des langages contrôlés existant à la MAE et une description de leur usage. Je tenterai de faire quelques rapprochements entre ces descripteurs et les mots-clés du JSA.

La seconde section sera consacrée à la recherche d'informations. Un bilan sera également dressé à partir des données recueillies pendant l'enquête. J'introduirai les notions de produits et de services pour essayer de trouver les moyens d'améliorer certes la recherche d'informations, mais aussi et surtout, pour rendre l'information facilement accessible et pour mieux la diffuser.

8 Indexation et mots-clés : quelques préconisations...

L'enquête analysée dans la partie précédente doit être complétée par une description, d'une part, des langages existants dans la MAE : les PACTOLS et Géoethno, d'autre part, par une mise à plat de quelques mots-clés d'auteurs et de revues. En effet, dans certaines revues comme *Paléorient*, les articles sont indexés avec d'autres mots-clés que ceux donnés par les auteurs et ceux des langages contrôlés.

Mais, avant d'aller plus loin, faisons un point sur les besoins en matière de fabrication de mots-clés.

8.1 Bilan des besoins pour la fabrication des mots-clés

À travers les entretiens réalisés, j'ai pu relever un certain nombre de points interprétables en termes de besoins et/ou de fonctions.

Le premier concerne le nombre de mots-clés. Plusieurs auteurs trouvent qu'il n'est pas assez important. Cinq-six, ce n'est pas suffisant pour rendre compte du contenu d'un texte. Il en faudrait facilement huit-neuf.

Le deuxième est relatif à des consignes au sujet des mots-clés. Très peu de revues imposent des règles quant à la constitution de mots-clés. Les auteurs se trouvent un peu démunis dès qu'il s'agit d'en fabriquer. En même temps, on peut se demander quel type de consignes pourrait donner une revue, dans la mesure où les textes sont très différents les uns des autres.

Le troisième est d'un niveau fonctionnel, il porte sur l'indexation proprement dite. Il est, en effet, apparu que les mots-clés d'auteurs ne servent pas à l'indexation des textes ou s'ils servent à cela, ils ne sont pas très efficaces car ils ne sont pas du tout contrôlés. La question est donc de savoir : ne serait-il pas intéressant de faire en sorte que ces mots-clés aident à l'indexation d'une façon ou d'une autre ?

De ces trois aspects, il semble découler au moins trois scénarii.

Le **premier scénario** serait d'établir une liste de mots prédéfinis, dans laquelle les auteurs pourraient venir piocher. Cela leur donnerait un cadre de référents et l'indexation serait un peu plus contrôlée. Cette liste serait fermée, son inconvénient serait donc que certains auteurs n'y trouvent pas les termes appropriés pour leur texte.

Un **deuxième scénario** serait de donner quelques pistes aux auteurs pour faire leurs mots-clés, de les conserver en considérant qu'ils sont la propriété de l'auteur et de s'en inspirer pour réaliser une indexation qui pourrait prendre appui sur un langage contrôlé déjà existant ou sur un vocabulaire établi par la revue. Par « pistes » aux auteurs, il faut entendre quelques orientations comme de mettre une aire géographique, un groupe ethnique, une période, un terme pour le thème général et deux pour les thèmes spécifiques.

Une **troisième scénario** serait un mélange des deux premiers. Une liste de mots-clés prédéfinis serait mise en place à l'intention des auteurs, dans laquelle ils pourraient choisir leurs mots-clés. Ils auraient également la possibilité d'en proposer des nouveaux si ceux de la liste ne conviennent pas. Ces termes nouveaux seraient acceptés ou non par la rédaction, donc ajoutés ou non à la liste. Cette liste serait donc contrôlée et validée par un comité d'experts. On peut imaginer qu'elle aurait sûrement un impact positif sur les recherches d'informations par mots-clés.

Liste de mots-clés et/ou vocabulaire d'une revue doivent être mis en regard avec les langages existants dans la MAE car, si on pouvait trouver le moyen de les rapprocher, on gagnerait en efficacité et en pertinence.

8.2 Deux thésaurus à la MAE

Un travail sur les langages documentaires a été réalisé au sein de la Maison René-Ginouvès par le pôle documentaire [33]. Je vais à présent décrire ces langages avant de regarder la façon dont ils sont utilisés par certains services.

8.2.1 Les PACTOLS

8.2.1.1 Description

Parmi les services et missions proposés par FRANTIQ [34], on trouve les thésaurus PACTOLS [35], acronyme de « Peuples et cultures, Anthroponymes, Chronologie relative, Toponymes, Oeuvres, Lieux, Sujets ». Chaque initiale (ex : P pour Peuples) correspond à un micro thésaurus sauf les Toponymes (liste alphabétique de termes).

33. Voir annexe n° 2.
34. Frantiq (Fédération et ressources sur l'Antiquité) est une coopérative de centres de recherches du CNRS, des universités et du Ministère de la culture, mettant en commun des bases de données sur les sciences de l'Antiquité, de la préhistoire au Moyen Âge.
35. Pour avoir une vision détaillée des PACTOLS, on pourra se reporter au mémoire de Loraine Marcheix [18] et au portail de Frantiq : http://frantiq.mom.fr/portal.php.

> « Peuples » regroupe tous les noms de cultures préhistoriques, les peuples antiques ou modernes et les regroupements humains tels que les habitants d'une cité et les ordres religieux.
>
> « Anthroponymes » regroupe tous les noms de personnes (dieux, héros, écrivains, gouvernants…) dont on parle dans un document. Les chiffres romains ont été transformés en chiffres arabes pour les dynasties et les gouvernants.
>
> « Chronologie » regroupe tous les termes de chronologie relative et les ères géologiques.
>
> « Toponymes » constitue une liste ouverte de lieux-dits, avec des règles d'écriture.
>
> « Œuvres » regroupe toutes les oeuvres artistiques et littéraires quelles soient religieuses, juridiques ou poétiques.
>
> « Lieux » regroupe tous les noms de lieux hiérarchisés à partir du continent jusqu'à la commune, la géographie physique et hydrographique des continents ainsi que des rubriques sur le monde antique.
>
> « Sujets » regroupe tous les mots-clés thématiques. [36]

Ce sont donc des listes de termes sélectionnés et organisés selon des relations sémantiques (synonymie, antonymie, adjacence) et/ou hiérarchiques (termes génériques et termes spécifiques). Ces thésaurus sont poly-hiérarchiques (un terme spécifique peut être relié à plusieurs termes génériques) et multilingue. La version actuelle comporte, d'une part, la traduction quasi intégrale des Sujets et de la Chronologie en italien et en anglais, d'autre part, la traduction de termes isolés traduits en allemand et en espagnol.

Les thésaurus PACTOLS sont évolutifs (mise à jour constante) et téléchargeables en rtf. Ils sont gérés par le logiciel libre Open Theso, avec une version publique Open Theso web et donc accessible pour tous. De plus, Open Theso est exportable en SKOS (Simple Knowledge Organisation System), donc interopérable [37]. En d'autres termes, il s'agit d'un langage permetant la concordance (ou le *mapping*) entre plusieurs langages documentaires

36. Toutes les informations mentionnées ici ont été constituées à partir de la présentation des PACTOLS faite sur le site : http://frantiq.mom.fr/html/pactols/pactols.html (consulté le 09 octobre 2009).
37. http://193.48.137.41/OpenThesoWeb/Navigation?action=retour_nav (consulté le 24 août 2009).



Fig. 13 – Copies d'écran la page des PACTOLS sur Open Theso web.

Les PACTOLS sont en accès libre et gratuit sur Internet. Comme on peut le voir, plus haut, sur la copie d'écran, il est possible de les consulter soit par navigation dans les micro-thésaurus, soit par recherche, permettant de choisir entre plusieurs modes d'affichage (listes hiérarchiques, listes alphabétiques, listes permutées). Enfin, des règles d'écriture régissent la création des concepts. Seuls ces termes peuvent être utilisés pour l'indexation et l'interrogation. Un lien avec une base de données peut être directement établi et permet de visualiser les données indexées avec le descripteur retenu.

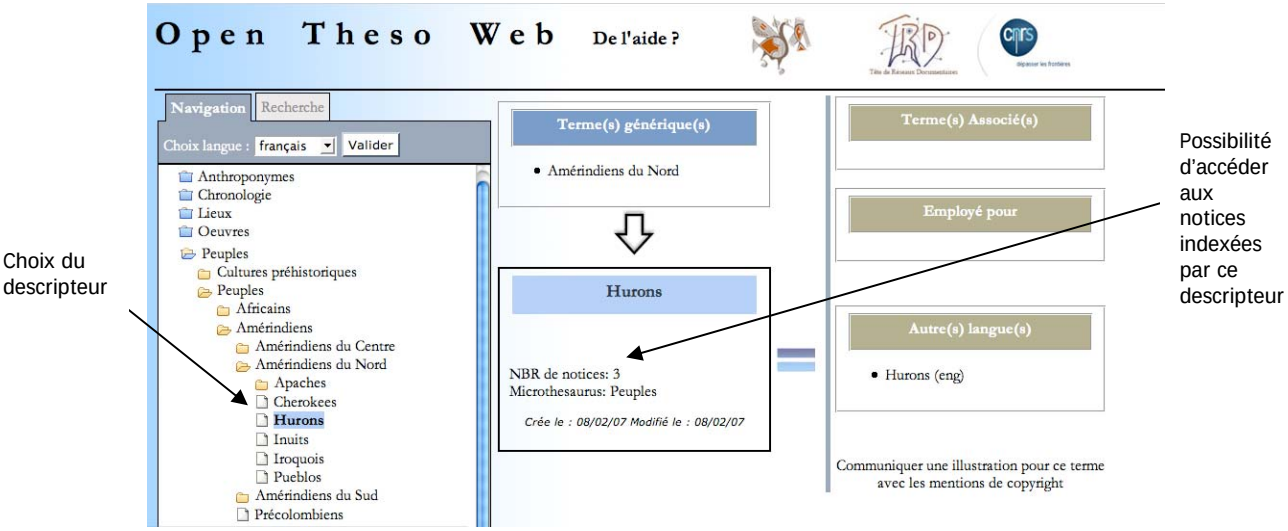


Fig. 14 – Copie d'écran pour l'accès aux descripteurs des PACTOLS.



Fig. 15 – Copies d'écran de la liste des notices indexées par un descripteur des PACTOLS.

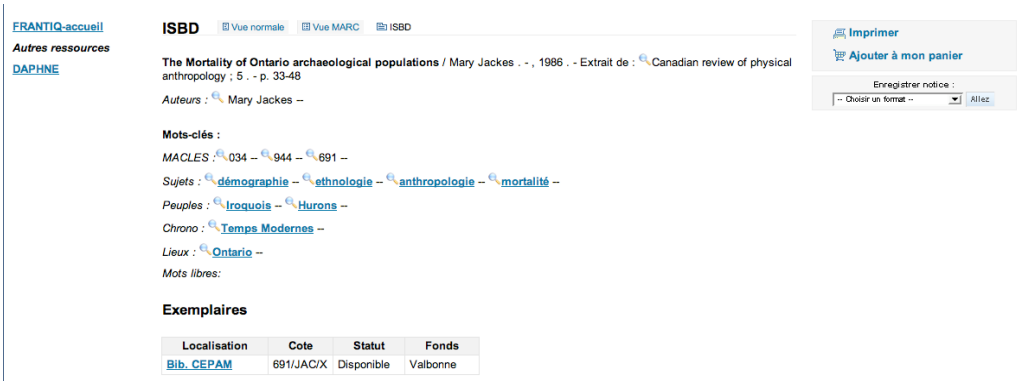


Fig. 16 – Copie d'écran d'une des notices indexées par le descripteur retenu.

8.2.1.2 Usages

D'après Marcheix ([18] Marcheix, p. 70), les PACTOLS sont utilisés par deux catégories de personnes, d'une part, des gestionnaires qui ont en charge de faire évoluer le thésaurus, d'autre part, « des utilisateurs qui se divisent eux-mêmes en deux catégories : les indexeurs et personnels des bibliothèques, et les étudiants et chercheurs – "le public" ».

Pour cette dernière catégorie, il faut considérer, d'un côté, ceux qui vont utiliser ces langages pour faire une recherche d'informations, c'est le cas des étudiants, des chercheurs…, de l'autre côté, ceux qui s'en servent pour faire de l'indexation des documents, il s'agit cette fois du personnel de la bibliothèque d'archéologie, du service des archives et d'AdlFI.

J'ai déjà montré dans la première partie en quoi consistait l'opération d'indexation. Rappelons qu'elle a pour but de décrire le contenu d'un document grâce à des concepts, des mots-clés, des termes résultant d'une analyse et pouvant provenir d'un langage contrôlé. En d'autres termes, on peut parler d'une sorte de traduction d'un contenu en concepts. Dans le

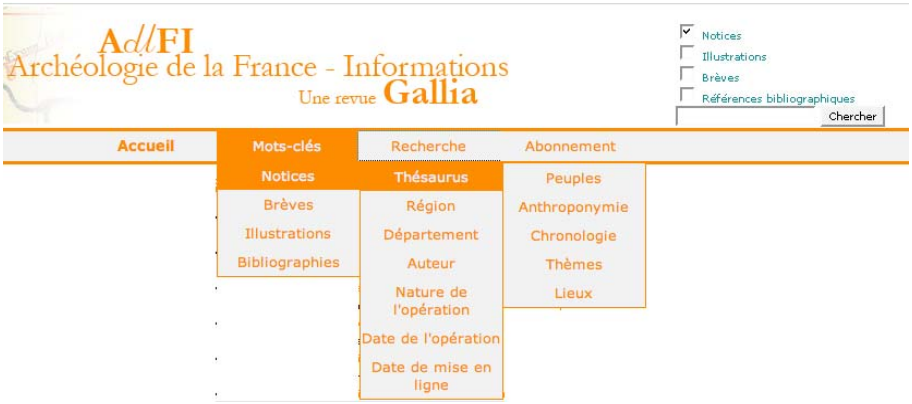
cas de ces services, cette traduction se fait par le biais des PACTOLS. Toujours selon Marcheix ([18] Marcheix, p. 71),

> L'indexation doit répondre à plusieurs questions qui trouvent leur réponse dans chacun des micro-thésaurus des PACTOLS :
> - Quoi ? Sujets
> - Où ? Lieux et/ou Toponymes
> - Quand ? Chronologie et/ou Cultures
> - Qui ? Anthroponymes ou Peuples
> - Comment ? Sujets ou Œuvres

L'indexation doit répondre à ces questions dans le but de rendre la recherche d'informations efficace. Cette fois, les utilisateurs de la première catégorie vont non pas indexer les documents recherchés, mais leur requête grâce à ces thésaurus.

Revenons sur AdlFI [38] (*Archéologie de la France – Informations*) qui est produit au sein du pôle éditorial. Il s'agit en effet d'une revue, coéditée par le Ministère de la culture (sous-direction de l'Archéologie) et le CNRS et destinée à publier l'actualité des opérations archéologiques menées en France. Les informations scientifiques préparées par les services régionaux de l'archéologie sont alors éditées et mises en ligne dans une base de données à l'usage des chercheurs nationaux et internationaux, des divers acteurs de l'archéologie et de tous ceux qui s'intéressent au patrimoine national. (*Archéologie de la France – Informations*)

Pour indexer les notices de cette revue, les secrétaires de rédaction utilisent, de manière autonome, les PACTOLS et peuvent être amenés à faire des propositions de termes si cela est nécessaire.



38. Voir http://www.adlfi.fr/SiteAdfi/index (consulté le 09 octobre 2009).



Fig. 17 – Copies d'écran d'ADLFI montrant la mention des PACTOLS.

Ce serait sans doute de cette façon que les PACTOLS pourraient être utilisés par d'autres revues d'archéologie pour faire une indexation contrôlée des textes publiés.

8.2.2 Géoethno

Il existe un autre thésaurus, produit à la MAE, par une documentaliste de la bibliothèque du laboratoire d'ethnologie et de sociologie comparative (LESC) : il s'agit de GeoEthno [39], thésaurus des noms géographiques. Il a été réalisé « pour remplacer celui fourni par la base Francis datant de 1983 et devenu obsolète » ([5] Donze p. 177).

8.2.2.1 Description

GeoEthno est un thésaurus géographique permettant l'indexation géographique de documents dans le domaine de l'ethnologie.

Il a été construit à partir d'un corpus de mots-clés géographiques, créés avec l'informatisation de la bibliothèque en 1985. Le corpus a dû être nettoyé et enrichi à la suite d'un travail sur des atlas et des listes de références. Simple liste de mots-clés à l'origine, il a été organisé en liste structurée, en s'appuyant sur la liste des noms de pays et territoires de l'ISO (ISO 3166-1:1997: Codes pour la représentation des noms de pays et de leurs subdivisions. Partie 1, Codes Pays – Liste en français) et sur le découpage des régions macro-géographiques de la Division des statistiques de l'ONU (Classification « m49 »).

La base de données GeoEthno n'est pas une base de données toponymiques, mais bien un thésaurus. Certains termes sont retenus comme descripteurs à des fins d'indexation dans le catalogue de la bibliothèque. Ceux qui ne sont pas retenus (donc les non-descripteurs) sont

39. Se reporter à l'article d'Isabelle Donze ([5] Donze) et au site du thésaurus : http://www.mae.u-paris10.fr/dbtw-wpd/bibliotheque/g/MON.html (consulté le 09 octobre 2009).

généralement des synonymes, des quasi-synonymes, des variantes, des termes translittérés, des termes traduits...

Ce thésaurus couvre toutes les aires géographiques, en prenant les grands continents comme termes génériques : Afrique, Asie, Europe, Océanie, Amérique.

Construit en SGML et accessible sur le web, il comporte 15 000 termes. Sa mise en place a donc dû répondre à un certain nombre de contraintes comme « l'intégration du corpus des 2700 mots-clés constitués depuis 1985 et disponible sous forme d'un index (liste à plat), possibilité de restructuration des données pour s'adapter à d'éventuels changements géopolitiques, facilité de saisie et de mise à jour, maintenance dans le temps, portabilité entre systèmes informatiques hétérogènes, consultation sur support imprimé ou électronique, travail collaboratif entre bibliothèque » ([5] Donze).

8.2.2.2 Usages

Ce thésaurus sert non seulement à l'indexation des documents, mais aussi à la recherche d'informations sur la base de données de la bibliothèque du LESC et, plus largement, sur celle du Réseau ethnologie [40].

Grâce à cette interface (voir copies d'écran), il est possible de consulter l'arborescence de ce thésaurus. Cela permet d'avoir une vue générale de la structuration des données et du degré de précision vers lequel ce thésaurus peut conduire.

40. Constitué officiellement en juin 2000, le Réseau Ethnologie est un réseau documentaire thématique et repose aujourd'hui sur la collaboration de quatre bibliothèques (Documentation du centre de recherches bretonne et celtique, Centre de documentation André-Georges Haudricourt, Bibliothèque du Laboratoire d'anthropologie sociale, Bibliothèque du Laboratoire d'ethnologie et de sociologie comparative). Il est principalement orienté sur la production du catalogue commun, base de données documentaire de plus de 130 000 notices bibliographiques.



Fig. 18 – Copie d'écran de la vue de l'arborescence du thésaurus GeoEthno.

Par l'arborescence, on ne peut visualiser que les termes descripteurs, les non-descripteurs n'apparaissent pas. On ne peut pas non plus accéder aux zones transversales. On peut donc choisir le (ou les) descripteur(s) pertinent(s) pour l'indexation du document et, du coup, le degré de précision pour cette indexation.

Le mode « interrogation » fonctionne à partir d'une base de données. En interrogeant un terme, on obtient sa localisation dans un pays ou une zone transversale.

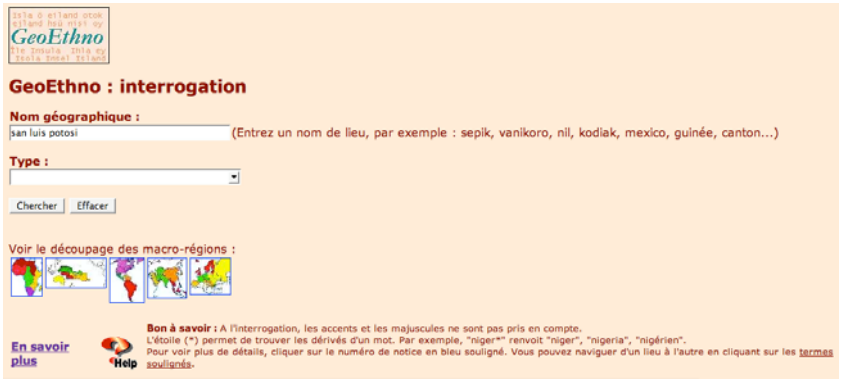


Fig. 19 – Copie d'écran de mode interrogation du thésaurus GeoEthno.

Par exemple, pour San Luis Potosi, on obtient comme première réponse :



Fig. 20 – Copie d'écran de la réponse obtenue dans le thésaurus GeoEthno.

En cliquant sur le chiffre 1, on obtient davantage d'informations :



Fig. 21 – Copie d'écran des informations complémentaires répondant à la requête précédente.

L'arborescence comme le mode « interrogation » facilitent donc la navigation dans le thésaurus.

Ce thésaurus s'adresse à deux catégories d'utilisateurs : d'une part, les documentalistes du réseau ethnologie qui font de l'indexation pour une base de données et, d'autre part, l'utilisateur comme un chercheur, un étudiant ou encore toute personne intéressée par l'ethnologie. Pour ce deuxième type d'utilisateur, le mode « interrogation » permet de mieux prendre en main le thésaurus et de pouvoir l'explorer facilement.

Pour avoir accès aux documents, il faut passer par le catalogue du réseau ethnologie, ainsi que le montrent les copies d'écran suivantes.



Fig. 22 – Copie d'écran de la page « interrogation avancée du catalogue », réseau ethnologie.



Fig. 23 – Copie d'écran des résultats obtenus.

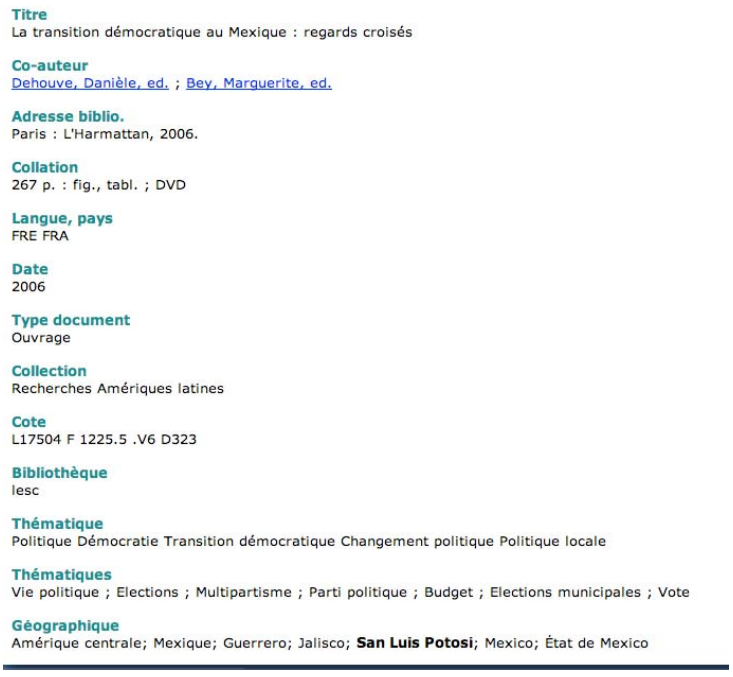


Fig. 24 – Copie d'écran de la notice complète du résultat 3.

8.3 Rapprochement mots-clés et descripteurs

Pour me rendre compte du contenu de ces langages et de la liste de mots-clés de certaines revues, essentiellement du JSA [41], j'ai tenté de faire une étude de certains mots-clés proposés dans l'index Géographique/ethnique du JSA. D'une part, je les ai mis en regard avec les descripteurs proposés dans Géoethno [42] pour voir s'il existe des concordances ou non, pour définir une façon d'indexer… D'autre part, j'ai mis à plat les descripteurs portant sur les peuples d'Amérique dans les PACTOLS, bien que l'ethnologie ne fasse pas partie du champ couvert par ce thésaurus, du moins pour le moment. Puis, j'ai comparé avec les mots de l'index Géographique/ethnique du JSA.

Au regard des exemples choisis, il apparaît que, grâce au thésaurus GeoEthno, on pourrait donner une indexation plus précise des documents et plus harmonieuse, ne serait-ce par exemple qu'au niveau de l'orthographe des noms de lieu pour éviter d'avoir des documents sur la même aire géographique indexés par deux termes différents parce que le nom de lieu est écrit de deux manières.

De plus, on pourrait se mettre d'accord sur la façon d'indexer, en allant du générique au particulier, et en évitant des expressions peu précises comme Amazonie du nord-ouest qui, contrairement à ce qu'elle paraît, définit une zone géographique, non pas un territoire proprement dit.

Cela permettrait de diminuer le bruit, mais surtout le silence dont il n'est pas toujours facile de se rendre compte, dans la recherche de documents

De plus, un certain nombre de termes géographiques que l'on trouve dans les textes publiés dans le JSA n'apparaissent pas dans GeoEthno, ne pourrait-on pas les proposer pour enrichir ce langage ?

De la même façon, j'ai regardé ce qui pouvait intéresser le JSA dans les PACTOLS et j'ai trouvé quelques descripteurs concernant les peuples comme on peut le voir sur la copie d'écran suivante.

---

41. Faute de temps et de compétences dans le domaine des autres revues, je n'ai pu me concentrer véritablement que sur le JSA. Il serait, pourtant, très intéressant de faire le même travail sur *Paléorient* et la *Revue archéologique*.
42. Ce travail ne se veut en aucun cas exhaustif, il s'agit de l'étude de quelques exemples de mots-clés pris dans le JSA. Le tableau mis en annexe n° 4 rend compte des résultats de cette analyse.

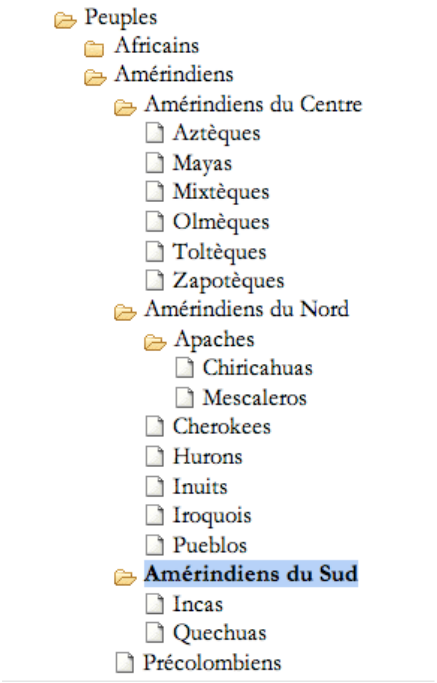


Fig. 25 – Copies d'écran des Peuples amérindiens dans les PACTOLS.

On constate qu'ils sont très peu nombreux et ne couvrent donc pas la richesse de l'Amérique. Dans le cas précis des PACTOLS, il semble nécessaire de proposer un travail de collaboration pour développer toute la partie « Amérique ». Celle-ci ne pouvait être développée auparavant dans la mesure où ces thésaurus n'avaient pas pour but premier de couvrir les thématiques américanistes, ni l'aire géographique, ni la chronologie.

*
* *

La description de ces différents thésaurus est une entrée en matière sur les langages existant à la MAE et qui pourraient être utilisés pour indexer les revues. En juste retour des choses, celles-ci pourraient, peut-être, aider à l'enrichissement de ces langages grâce aux mots-clés des auteurs.

9 Recherche d'informations, diffusion et/ou signalement d'informations

Par recherche d'informations, on a coutume d'entendre méthodes et outils permettant de trouver l'information pertinente répondant à une requête. Je propose ici de regarder aussi la recherche d'informations du point de vue de la diffusion, c'est-à-dire à travers les méthodes et outils permettant de faire arriver l'information jusqu'à la personne pouvant être intéressée par celle-ci.

9.1 Bilan des besoins en termes de recherche d'informations

D'après l'enquête menée auprès d'un nombre restreint de chercheurs, on constate que trois sont souvent satisfaits du résultat de leur recherche [43] ; trois autres ne le sont pas toujours, voire jamais ; une dernière personne est très satisfaite : il s'agit du chercheur qui a mis en place un système de veille, notamment pour tout ce qui concerne les publications à l'étranger. Elle aimerait en avoir plus, c'est-à-dire avoir accès à d'autres ressources documentaires, pas toujours accessibles sur Internet et/ou pas assez diffusées.

On peut avancer, en prenant bien garde ne pas en faire une généralité étant donné le nombre de personnes interrogées, que plus de la moitié des personnes interrogées est *grosso modo* satisfaite du résultat de leur recherche d'informations. Pourtant, derrière cette satisfaction, se cachent quelques plaintes.

Plusieurs se plaignent en effet du bruit dans les réponses obtenues : s'ils peuvent le considérer comme normal sur Google ou Google Scholar, ils le trouvent particulièrement énervant sur des sites spécialisés (par exemple, sur Jstor) car ils considèrent que tout doit être important puisqu'il s'agit d'un site spécialisé. Sur Google ou Google Scholar, ils regardent les premières pages seulement. Quand ils ont peu de réponses, ils pestent contre le système et ne comprennent pas vraiment pourquoi ils n'ont pas plus de réponses, notamment sur un site spécialisé. Le silence peut donc surprendre en ce sens qu'il est

43. Une enquête auprès d'un plus grand nombre de personnes a été menée en 2005-2206, dont les résultats sont très intéressants et rejoignent les préoccupations mentionnées dans cette partie. Se reporter au Powerpoint réalisé pour « "L'atelier documentaire du chercheur" : de nouveaux enjeux pour l'IST à l'INRIA », par Laurent Bouvier-Ajam et Kumar Guha, sur le site des RPIST 2009 (http://rpist.inist.fr/spip.php?rubrique54, consulté le 26 septembre 2009).

étrange de ne pas trouver de documents sur le sujet interrogé ; ils ne remettent pas en question leur manière de l'interroger.

Certains montrent un peu de méfiance vis-à-vis des résultats, surtout sur Google, car il n'est pas toujours possible et/ou facile de vérifier l'information, sa validité, sa provenance, son auteur. Sur les sites spécialisés, ce problème ne se pose pas vraiment, du moins sur le site des revues puisque les textes publiés ont été validés par les pairs.

Une des personnes rencontrées dit qu'elle ne pourra jamais être satisfaite car, selon elle, « ce qui est le plus désagréable, c'est qu'avec le développement d'Internet on a l'impression qu'on peut avoir accès à tout, or ce n'est pas vrai ». Avec le développement d'Internet, des frustrations se créent quand on ne peut pas avoir accès au document. Il devient difficile de se satisfaire d'une simple localisation physique.

Deux autres personnes interrogées évoquent les difficultés à se retrouver devant tous ces portails, ces outils existants… L'une d'elle dit qu'elle a l'impression de « bidouiller ». L'autre justifie son utilisation de Google à cause de cette multitude et de cette diversité de ressources, pas toujours évidentes à connaître et à interroger.

Enfin la plupart reconnaissent que la recherche proprement dite d'informations prend beaucoup de temps, qu'ils aimeraient pouvoir dépouiller régulièrement les revues de leur domaine, lire les livres parus ou au moins les feuilleter, assister à toutes sortes de manifestations (colloques, séminaires, conférences congrès)… Mais le temps manque considérablement.

Revenons sur les trois types de recherche que j'ai tenté de faire ressortir dans la partie précédente car, d'une façon ou d'une autre, elles permettent de resserrer les besoins en matière de recherche d'informations, mais aussi de donner des pistes pour répondre à ces plaintes :

-une **recherche de type ponctuel** qui répond à un besoin limité dans le temps (pour la préparation d'un cours, d'une conférence, d'un exposé…).

-une **recherche de type continu** qui répond à un besoin d'être tenu régulièrement informé sur un sujet précis, pas nécessairement central, parfois annexe au thème d'étude (pour connaître l'actualité d'une discipline ou d'un domaine, une découverte…).

-une **recherche de type courant** qui répond à un besoin de suivi des parutions et/ou manifestations concernant son propre domaine d'étude, sa spécialisation.

Si le premier type défini correspond bien au fait de trouver l'information qui convient, les deux autres types demandent plus de temps et d'endurance car ces recherches se font sur du long terme et de façon régulière. Pour ces trois types de recherche, il existe des réponses adaptées, des réponses d'autant plus adaptées qu'on peut les circonscrire dans le cadre d'une revue spécialisée et/ou d'un pôle de revues spécialisées.

9.2 Des propositions de développement et/ou d'améliorations

J'ai choisi de placer l'indexation au centre de ma réflexion parce qu'il me semble indéniable qu'une bonne indexation des documents va permettre d'orienter la recherche d'informations dans ce labyrinthe foisonnant de documents numériques et de développer des outils et des services utiles pour les chercheurs et étudiants en ethnologie et archéologie.

Que doit-on entendre par une « bonne » indexation ? On peut répondre à cette question en imaginant idéalement qu'une indexation pourrait être considérée comme bonne si les termes utilisés pour rendre compte du contenu d'un document correspondent exactement à ceux utilisés pour définir le contenu d'une requête : une concordance parfaite ! Réaliser une « bonne » indexation n'est donc pas simple, car on peut difficilement prévoir exactement et avec certitude la façon dont les personnes vont interroger une ressource documentaire. Néanmoins, de façon plus réaliste, on peut considérer qu'une « bonne » indexation est ce qui va permettre de donner des résultats dont la pertinence sera au plus près des termes de la requête.

De plus, pour trouver une parade à la difficulté de ne pas savoir comment les personnes vont s'y prendre pour rechercher de l'information, on peut peut-être les amener à interroger les ressources en suivant une procédure, pas trop lourde, en leur fournissant des consignes d'interrogation, des suggestions de recherche, bref en les orientant dans leur recherche grâce à des outils comme un plan de classement, des mots d'interrogation… Cet aspect ne pourra réellement être développé que si l'information est bien structurée en amont.

Posons le problème autrement. La mise en ligne des revues est sans aucun doute une très bonne chose puisque, d'une part, elle permet aux personnes intéressées d'avoir accès aux textes ou de connaître le sommaire des derniers numéros parus, d'autre part, elle permet de faire connaître la revue, de la rendre visible et de la diffuser plus largement. Néanmoins, juste une mise en ligne ne suffit pas car la recherche plein texte n'est pas une solution toujours satisfaisante quand on veut trouver des informations générales, par exemple sur un thème.

De plus les revues scientifiques sont par définition spécialisées, voire très spécialisées. De ce fait, la recherche d'informations est non seulement très ciblée, mais aussi très circonscrite par le domaine couvert par la revue. Qui plus est, les informations publiées sont assurées d'une reconnaissance scientifique puisqu'elles ont été approuvées par les pairs avant de pouvoir être publiées. Ces aspects de spécialisation, de circonscription des sujets en raison de cette spécialisation et de validité scientifique des informations publiées, vont de pair avec l'idée d'une structuration de l'information. De par son organisation interne, j'ai posé qu'une revue scientifique pouvait être considérée comme un système d'informations organisées.

Avec la mise en ligne sur Revues.org, les textes publiés dans le JSA sont structurés grâce notamment à l'utilisation des métadonnées Dublin Core. Mais là où le bas blesse, c'est sur le contenu proprement dit. L'indexation du contenu laisse à désirer alors qu'étant donné la spécialisation des revues scientifiques il serait possible de remédier à ce mal. Cela ne pourra véritablement se faire de façon efficace que grâce à la création et/ou à l'utilisation de langages contrôlés [44]. En effet, l'usage de ces langages pour les textes publiés dans une revue pourrait rendre l'accès à l'information meilleur et permettrait d'envisager le développement de produits et de services tels que des systèmes de signalement de l'information et/ou de diffusion de l'information présélectionnée (par exemple grâce à des flux RSS, des alertes, des profils utilisateurs).

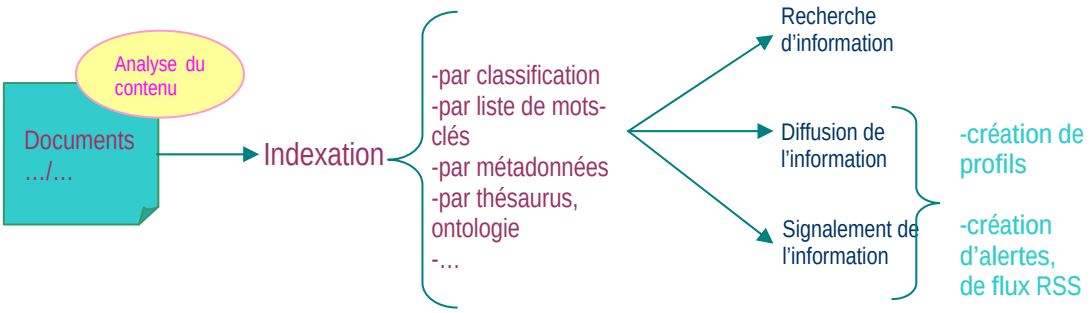


Fig. 26 – Schéma de valorisation de l'information grâce à l'indexation.

Pour arriver à un tel résultat, il est nécessaire de réfléchir à des outils, somme toute assez classiques, comme des bases de données d'une part et des sites web d'autre part.

9.2.1 Bases de données

Dans certaines revues du pôle, il existe déjà des bases de données. La revue *Paléorient*, par exemple, en propose deux : l'une consacrée aux articles parus dans la revues ; l'autre réservée aux nouvelles références citées dans les bibliographies d'articles. Cette dernière est

44. Bien sûr, il ne s'agit pas seulement de thésaurus et/ou d'ontologie. Une classification pourra être aussi utilisée.

donc semblable à une bibliographie spécialisée sur un domaine. La première rappelle les tables et/ou index papiers de revues.

Au sein du pôle éditorial de la MAE, une réflexion commune serait intéressante à mener à propos du développement de base de données car, même si les revues sont très différentes au niveau du domaine traité, du statut administratif et des ressources financières, leurs besoins en matière de gestion, de diffusion et de visibilité de l'information restent, à peu de choses près, semblables.

9.2.1.1 Bases de données de textes publiés

Le besoin s'est fait sentir dans plusieurs revues d'avoir une base de données des textes parus. Elle pourrait remplacer la version papier des tables et/ou index qui était (qui est encore, pour certaines) publiée à périodicité variable selon la revue. Mieux que ces outils papiers, elle permettrait de faire des recherches sur l'ensemble des numéros de la revue, des recherches sans doute plus performantes que celles réalisées grâce au moteur proposé sur les plateformes de mise en ligne. Il s'agirait nécessairement de bases de type relationnel dans la mesure où un texte peut être écrit par un ou plusieurs auteurs et qu'un auteur peut écrire un ou plusieurs textes.

La structure de ces bases de données pourrait être composée de plusieurs champs, assez courant dans une revue scientifique : auteur(s), titre, année, numéro de la revue, page, url du document numérique avec accès directement au texte quand celui-ci est en ligne, résumés et mots-clés d'auteurs, thèmes généraux, thèmes spécifiques, régions/aires géographiques, peuples, chronologie. Le recours à un vocabulaire contrôlé prend ici toute sa place. À cette première série de champs pourraient s'ajouter des champs concernant les auteurs (avec leurs coordonnées personnelles et institutionnelles, domaine de recherche, thèmes spécifiques…). Cette deuxième série de champ devra être confidentielle et réservée aux personnes gestionnaires de la base (les secrétaires de rédaction, par exemple). Cela permettra d'accroître le carnet d'adresses de la revue tout en fournissant des détails importants pour des éventuelles collaborations autres avec ces mêmes auteurs.

La première série de champs sera, elle, en revanche consultable et interrogeable par les utilisateurs. Il sera, en effet, possible d'interroger sur tous les champs et/ou sur un champ particulier. Une aide devra être proposée notamment pour tout ce qui touche au vocabulaire contrôlé. Des possibilités de se créer un profil et/ou une alerte devront être envisagées.

Enfin il faudra prévoir une promotion de ces bases pour les faire connaître et accroître leur utilisation.

9.2.1.2 Bases de données pour des ouvrages reçus et/ou parus et des comptes rendus de lecture

Dans un esprit assez proche, il pourrait être développé des bases de données sur les ouvrages reçus et/ou parus dans le domaine. Cela rejoint d'une certaine façon la base de données bibliographiques de *Paléorient*, excepté que l'objectif de la base que je propose est un peu différent. L'idée de cette base serait de motiver la rédaction de comptes rendus afin de dynamiser cette rubrique dans les revues. Il s'agit en effet d'une rubrique très importante. Les auteurs lisent volontiers les comptes rendus d'ouvrages car ce sont des textes courts et informatifs. Ils permettent de rester informé des nouveautés dans un domaine. De plus, ils sont souvent en ligne intégralement dès leur parution.

Ce type de bases de données pourrait avoir deux versants comme dans le type de base précédent : un versant utilisateur grâce auquel il serait possible de consulter la liste des livres reçus et/ou parus dans le domaine, de savoir si tel livre a déjà été pris pour compte rendu, si oui d'avoir accès au compte rendu, si non de pouvoir proposer d'en faire un. L'autre versant serait celui gestionnaire qui, comme dans les bases d'articles, reprendrait les informations sur les auteurs. Il est évident que ces deux tables d'auteurs, à la fois dans la base des articles et dans celles des comptes rendus, pourraient n'en faire qu'une car finalement ce serait un table des spécialistes du domaine. Elle permettrait, à la rédaction de la revue, de chercher éventuellement un spécialiste pouvant rendre compte d'un ouvrage.

*
* *

Quoi qu'il en soit, pour l'élaboration de ces deux types de bases, il semble évident qu'une collaboration entre les informaticiens, les documentalistes de la MAE et les secrétaires de rédaction du pôle faciliterait cette réalisation. En effet, il est important de réfléchir à une structure commune pour les bases de données des revues du pôle car c'est, je crois, le seul moyen de pouvoir, ensuite, proposer un moteur de recherche commun sur l'ensemble de ces bases. Qui plus est, si l'indexation pour chaque revue utilisait un vocabulaire contrôlé, c'est-à-dire précis, homogénéisé et harmonisé, la recherche en serait plus performante. Il serait possible alors de proposer d'autres types de services comme celui de créer une diffusion sélective de l'information (DSI) en fonction de profils que les utilisateurs pourraient créer. Cette DSI pourrait prendre la forme d'alerte, de flux RSS (voir plus loin)…

La première étape pour réaliser un tel projet serait la rédaction d'un cahier des charges à propos de ces bases de données, après avoir défini précisément les besoins de chacune des revues.

Pour que ces bases soient consultables via Internet, pour que ces projets se développent véritablement, il serait tout aussi important de réfléchir à des sites web propres à chacune des revues, mais aussi pour le pôle dans son ensemble.

9.2.2 Sites web

Encore une fois, la mise en ligne des revues ne remplace pas un site propre à chaque revue. Au sein du pôle éditorial de la MAE, quelques revues ont déjà leur site, même si les secrétaires de rédaction et directeurs de ces revues affirment la nécessité de le rafraîchir et/ou de le faire évoluer. Quoi qu'il en soit, l'idée de réaliser un site web pour chaque revue et pour le pôle est revenue dans plusieurs entretiens avec les SR et/ou directeurs de revues, concernant la valorisation des revues. Après leur mise en ligne (qui correspond au premier aspect à développer pour les rendre visibles, les valoriser et les faire connaître plus largement), il faut penser soit à faire un site web, soit à revoir celui qui existe déjà. Dans un tel contexte, il pourrait être intéressant de mener une réflexion commune pour la création d'un site web pour l'ensemble du pôle et pour chacune des revues.

Pour chaque revue, le site contiendrait évidemment des liens vers l'édition électronique et hébergerait les bases de données mentionnées plus haut. Des informations complémentaires pourraient être fournies sur les numéros à venir, sur les activités des associations éditrices de certaines revues, sur des événements en lien avec le domaine de la revue…

La difficulté aujourd'hui est de faire face à la multitude de sites et/ou de portails déjà existants et développés par les centres de recherches pour proposer toutes sortes d'informations concernant leur spécialité. Cette multiplicité contribue à l'éparpillement des informations. Il serait pourtant intéressant de trouver un moyen pour réunir toutes ces informations en un même endroit virtuel. Il est clair néanmoins que ce n'est pas facile à mettre en place. Cela poserait à terme des problèmes de mise à jour, demandant sans doute de mettre sur pied un système de veille qui, à la longue, pourrait occuper une personne à temps plein.

Pourtant, en raison du fait qu'une revue scientifique est en soi un repère pour les spécialistes du domaine qu'elle couvre, on aurait envie d'imaginer d'autres services sur son site, de façon à ce que celui-ci devienne, en quelque sorte, une référence dans le domaine de la revue. Des informations d'actualité (colloques, conférences, séminaires, expositions…) permettraient aux chercheurs et aux étudiants de connaître rapidement et facilement les manifestations qui ont lieu dans leur domaine, voire dans leur spécialité. Il est vrai que

Calenda [45] le propose, mais sur le domaine entier des SHS, ce qui ne facilite pas la récupération et la diffusion des informations.

De plus, sur le site de chaque revue, une liste d'adresses des lieux (bibliothèques, musées, centres de recherche, d'autres revues du domaine…) et/ou de sites ouvrirait les échanges entre personnes, mais aussi entre personnes et organismes, et entre organismes. De plus, une possibilité pourrait être offerte aux utilisateurs de faire des propositions pour compléter ces référencements.

Il serait dès lors possible d'envisager, sur ces sites, des pages d'onglets (comme ce que proposent les agrégateurs de flux RSS tels que Netvibes) organisés en fonction des thèmes et/ou des types d'informations retenues.

Enfin, rappelons qu'avec la mise en ligne des bases de données sur le site des revues, il sera possible d'envisager une diffusion sélective d'information. Elle prendra la forme de profil et/ou d'alerte et pourra se faire grâce à des termes sélectionnés par les usagers en fonction de leur centre d'intérêt. Ils seront dès lors avertis par courrier électronique que, sur le site de telle revue, un nouveau texte vient de paraître et peut les intéresser. Ils devront se rendre sur le site pour le consulter ou consulter son résumé. Une autre solution sera que les informations à propos de la parution d'un nouveau texte sur leur domaine arriveront directement dans leur boîte mail ou sur un système de veille (mis en place, par exemple, sur des agrégateurs de flux RSS) avec un lien (si celui-ci existe bien sûr) vers le document pouvant les intéresser.

*

Le site du pôle proprement dit proposerait, évidemment, des liens vers chacun des sites de revue. Il donnerait l'actualité du pôle, les activités mises en place, les différents événements programmés. Un espace de travail pourrait être réservé aux personnes travaillant dans les revues (SR, rédacteurs et/ou directeurs), sur lequel elles pourraient déposer des outils utiles à toutes les revues, préparer des réunions… Cela pourrait prendre la forme d'un espace de travail collaboratif.

*

* *

Dans tous les cas, là aussi, la première étape pour réaliser un tel projet serait la rédaction d'un cahier des charges à propos de ces sites, après avoir fait un bilan de ce qui existe déjà et après avoir défini précisément les besoins de chacune des revues et du pôle.

---

45. Calenda (http://calenda.revues.org/) est le plus important calendrier scientifique francophone en sciences humaines et sociales. Il est édité par Revues.org depuis 2000 et diffuse le programme de 10433 événements scientifiques. Il est possible de diffuser une information dans Calenda, en utilisant le formulaire de suggestion.

Conclusion

Partir des mots-clés et de l'indexation pour arriver à la valorisation de revues scientifiques n'est certes pas le chemin le plus direct, ni le plus court. Pourtant, il me semble difficile de nier qu'en matière de valorisation des revues, les mots-clés et l'indexation des revues et des informations qu'elles contiennent sont indispensables, pour ne pas dire essentielles, pour envisager une politique de valorisation homogène à un ensemble de revues spécialisées en sciences humaines et sociales.

Les revues peuvent se targuer d'être des réservoirs à mots-clés puisque les auteurs, donc des spécialistes reconnus par leurs pairs, fournissent des mots-clés. Ces derniers sont la propriété de l'auteur. Ils ne sont pas vraiment harmonisés d'un auteur à l'autre car aucune règle n'est donnée par les rédactions de revues pour leur fabrication. Ils ne sont pas non plus toujours utilisables pour l'indexation du texte, mais ils peuvent être une source d'information sur l'évolution du langage d'une discipline ou une indication pour trouver le bon mot-clé et/ou descripteur. Il faut donc bien distinguer mots-clés d'auteurs et mots-clés pour indexation, même si les deux peuvent se recouvrir, même si les premiers peuvent inspirer et/ou enrichir les seconds. Pourtant, étant donnée cette situation, il pourrait être intéressant de fournir aux auteurs des indications pour faire en sorte que les deux types de mots-clés se rapprochent le plus possible. C'est une réflexion qu'il faudra mener avec les secrétaires de rédaction et rédacteur et/ou directeur de revue intéressés par ce sujet.

De plus, pour arriver à une indexation correcte et surtout cohérente, nous avons montré qu'il pouvait être également intéressant d'utiliser un cadre, qu'il prenne la forme d'un langage documentaire déjà existant ou celle d'une liste de mots-clés plus ou moins contrôlés ou encore une autre forme. La forme de ce cadre devra être, elle aussi, décidée au sein même des rédactions de revues et, tant que faire se peut, en accord avec les professionnels de l'information qui font partie du pôle documentaire de la MAE et qui travaillent déjà sur et avec des langages documentaires. Pour le dire autrement, l'indexation est ce qui va, en quelque sorte, permettre de donner une structuration au contenu logique d'un document, de la même façon que l'attribution de métadonnées va permettre de structurer le support et le contenu physique de ce même document.

Le schéma qui suit permet de synthétiser plusieurs idées sur le rapport entre mots-clés, descripteurs et métadonnées et, donc, sur la structuration de l'information. Notons en effet que les mots-clés d'auteurs sont des mots sélectionnés par rapport à des thèmes de recherche ou à l'accueil que ces auteurs souhaitent pour leur article (ce sont donc des mots-clés « égocentrés »). Les mots-clés (ou descripteurs) des documentalistes sont, en revanche, des mots représentant un concept et sélectionnés par rapport à une discipline qui est décrite, grâce à un langage, en fonction de ses méthodes et contenus (il s'agit donc de mots-clés plus précis tout en étant plus ouverts et plus généralistes).

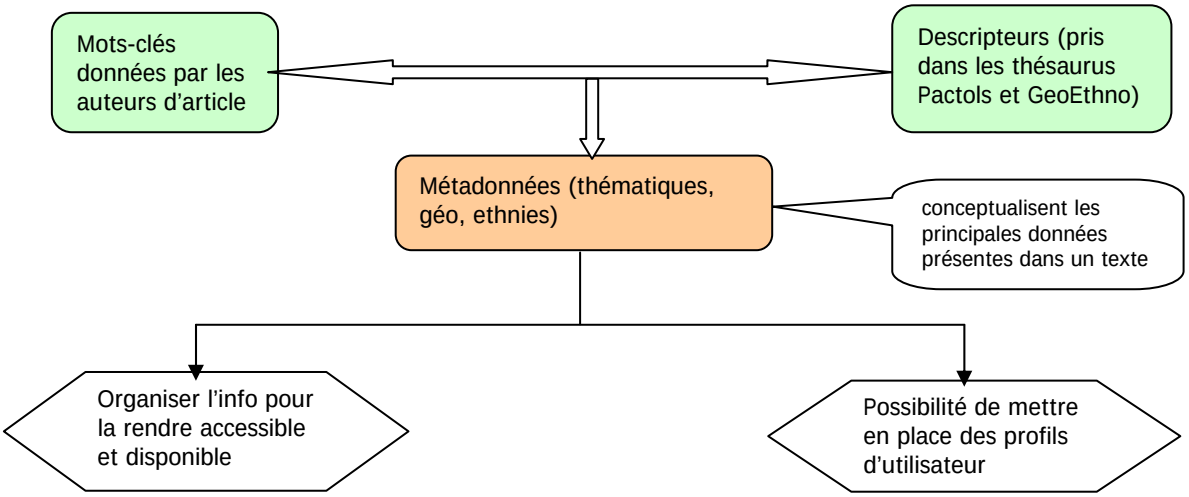


Fig. 27 – Schéma de mise en relation mots-clés d'auteurs, descripteurs et métadonnées.

Grâce à cette structuration, on pourra espérer que la recherche d'informations donne des résultats pertinents par rapport aux termes de la requête. C'est peut-être une illusion, mais il me semble plus facile de structurer l'information dans une revue puisque celle-ci est limitée par le domaine et/ou la spécialisation qu'elle couvre.

Plus encore que la recherche d'informations, proprement dite, cette structuration va aussi rendre possible une valorisation des informations contenues dans des revues scientifiques, puis l'ensemble du pôle éditorial. Cette valorisation prendra la forme d'outils, somme toute, assez classiques comme des bases de données et des sites web. Ces outils remplaceront ceux qui existaient en papier, mais permettront d'aller plus loin que ces outils papier en proposant des services à l'usager. Celui-ci pourra déposer son profil de recherche grâce à des mots-clés, des thèmes, des notions, des domaines correspondant à ces centres d'intérêt scientifique. Ces services pourront alors lui signaler et/ou lui diffuser toutes les informations et/ou les textes répondant à son profil.

Ces projets ne pourront être véritablement pensés et réalisés qu'avec l'aide et la collaboration des professionnels travaillant dans le pôle documentaire, d'une part, et des professionnels travaillant dans le service informatique, d'autre part.

Il existe déjà un projet collectif allant dans ce sens. Il s'agit de la mise en place d'un système intégré d'informations archéologiques multimédia. Dans ce premier intitulé, il manque le mot ethnologie, mais cela devrait être réparé très rapidement.

Dans le texte du quadriennal de la Maison René-Ginouvès archéologie et ethnologie 2009-2012, un paragraphe avait été rédigé sur ce projet :

Le bilan du pôle documentaire montre les richesses de chacun des services pour traiter les différents supports. Cependant un **accès conjoint** aux informations optimiserait les recherches. Il est donc important d'élaborer une recherche fédérée entre les services, et de prévoir une extension aux différentes équipes hébergées dans la Maison.

2.3.2.1. Conception d'un SIAM (Système intégré d'informations archéologiques multimédia)

L'objectif est de fournir au chercheur, depuis le site Internet, une page offrant un accès unique à tous les supports traités au sein de la Maison : fonds d'archives, monographies, articles et leurs textes intégraux, littérature grise, analyses documentaires, cartes et plans, photographies, etc. Un projet de création d'un méta-moteur de recherche réunissant les différentes bases et sites de la Maison avait été créé, mais n'ayant de crédits spécifiques pour ce projet, il a dû être réalisé avec des moyens internes.

Avec les évolutions technologiques récentes, il devient indispensable de mettre en place un moteur de recherche incluant de nouvelles fonctionnalités pouvant interroger les différentes bases et différents sites de la Maison. Une étude sur toutes les expériences réalisées ou initiées doit être menée afin d'établir une programmation, les outils adaptés pour intégrer tous les supports et les services de la Maison. Cette recherche fédérée utilisera la transformation des PACTOLS en ontologie pour l'archéologie et un système de géo-référencement des données.

Ce SIAM nécessite la collaboration effective du service Système d'Information (web et informatique) avec une prospective sur le web personnalisé (web 2.0). Il est donc prévu de mettre en place une démarche projet, avec l'élaboration d'un cahier des charges, une étude de faisabilité par un prestataire, un montage de projet, avec planification. Il est donc devenu nécessaire d'identifier les tendances-clés et les changements apparus lors du développement des modifications technologiques avec une attention particulière pour les nouveaux rôles, de même que pour les compétences et les habiletés nouvelles… (extrait du rapport quadriennal de la Maison René-Ginouvès – Unité mixte de service UMS 844 : soutien à une Maison des Sciences de l'Homme Contractualisation – 2009. Université Paris X-Nanterre, pagination multiple, pp. 48-49)

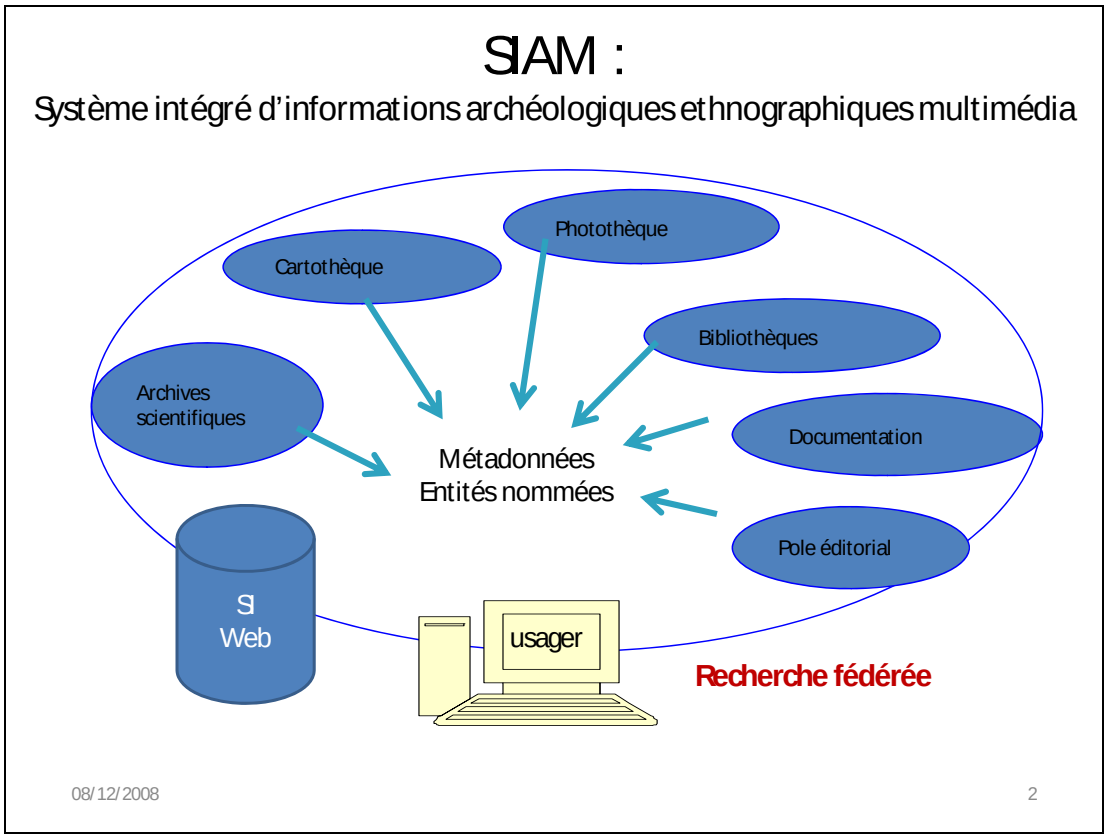


Fig. 28 – Schéma du projet Siam (réalisé par Brigitte Lequeux).

Une première réunion a eu lieu en décembre 2008 et avait pour but de réunir la plupart des protagonistes susceptibles de participer au projet SIAM. En sortant de cette réunion, il est apparu clairement que, pour que les revues puissent participer pleinement à ce projet car il répond précisément à certaines attentes en matière de valorisation des revues, il est nécessaire d'avoir autre chose à proposer qu'une simple mise en ligne, même si celle-ci représente déjà beaucoup de choses. Les sites web, peut-être, mais surtout, les bases de données mentionnées dans ce travail devront, en effet, être construits avant la réalisation effective de ce projet.

Pour finir, j'aimerais insister sur la nécessité de faire travailler ensemble des professionnels aux compétences complémentaires (les informaticiens, les documentalistes et les secrétaires de rédaction). Je ne peux m'empêcher de m'étonner face à l'absence de collaboration, notamment entre documentalistes et secrétaires de rédaction, surtout dans le domaine de la recherche scientifique, car il me semble que chacun de ces métiers apporte sa pierre à l'édifice, mais cet édifice ne pourra résister et se déployer que si chaque pierre est posée en tenant compte de celle d'à côté, en collaborant avec celle d'à côté...

Bibliographie

Dans cette bibliographie, les références sont classées selon six thèmes : Terminologie ; Indexation et langages documentaires ; Histoire des langages documentaires ; Recherche d'information ; Valorisation de l'information et web ; Édition papier/édition électronique. Les références citées dans ce mémoire sont numérotées pour faciliter les appels dans le texte. Des références complémentaires ont été ajoutées : elles ne sont pas citées, mais ont permis de nourrir la réflexion.

TERMINOLOGIE

[1] BOULOGNE Arlette. *Vocabulaire de la documentation*. Paris, ADBS Editions, 2004.

Références complémentaires

CACALY Serge (dir.). *Dictionnaire encyclopédique de l'information et de la documentation*. Paris, Nathan, 1997.

INDEXATION ET LANGAGES DOCUMENTAIRES

[2] CHAUDIRON Stéphane. « Technologies linguistiques et modes de représentation de l'information textuelle », *Documentaliste – Sciences de l'information*, 2007, vol. 44, n° 1, pp. 30-39.

[3] DESFRICHES DORIA Orélie. *La recherche d'information du grand public. L'évolution des langages documentaires ou l'avenir des ontologies : création d'un thésaurus dynamique pour le site de l'assitance en ligne de France Télécom Orange*. Mémoire pour obtenir le Titre professionnel « Chef de projet en ingénierie documentaire » INTD niveau I, Paris, INTD/CNAM, 2008.

[4] *Documentaliste – Sciences de l'information*. *Dossier Langages documentaires et outils linguistiques.* 2007, vol. 44, n° 1

[5] DONZE Isabelle. « "Pékin" ou "Beijing" ? La construction d'un thésaurus de nos de lieux et la problématique de l'indexation des noms géograhiques ». *In* Timimi Ismaïl et Susan Kovacs (coord.). *Indice, index, indexation. Actes du colloque international organisé les 3 et 4 novembre 2005 à l'Université de Lille-3 par les laboratoires Cersates et Gerico*. Paris/Lille, ADBS éditions/Université de Charles-de-Gaulle Lille-3/Gérico, 2006, pp. 177-190.

[6] KHAYARI Majid. « TermSciences : un portail pour valorsier la production terminologique des organismes publics de recherche ». *Arabesques*, 2009, avril-mai-juin, n° 54, pp. 16-17.

[7] Livre Blanc. *Valorisation de l'information non structurée*, Apil/Aproged/Cigref. Oct. 2007.

[8] MANIEZ Jacques. « Rupture ou continuité ». *Documentaliste – Sciences de l'information*, 2007, vol. 44, n° 1, pp. 12-16.

[9] RAÏS Nadia. *Principes et techniques d'indexation et de recherche de l'information textuelle : note technique*. Paris, INTD, juin 2009. 34 p. [support de cours]

[10] SERRE Alexandre. *Introduction à l'indexation*. URFIST Bretagne-Pays de Loire, Janvier 2002.
http://www.sites.univ-rennes2.fr/urfist/Supports/Indexation/PlanCoursIndexation.html (date de création : 15 janvier 2002, date de modification : 1 septembre 2003, date de consultation : 4 août 2009).

[11] TIMIMI Ismaïl et Susan KOVACS (coord.). *Indice, index, indexation. Actes du colloque international organisé les 3 et 4 novembre 2005 à l'Université de Lille-3 par les laboratoires Cersates et Gerico*. Paris/Lille, ADBS éditions/Université de Charles-de-Gaulle Lille-3/Gérico, 2006.

HISTOIRE DES LANGAGES DOCUMENTAIRES

[12] MANIEZ Jacques. *Actualité des langages documentaires, Les fondements théoriques de la recherche d'information*. Paris, ADBS Editions, 2002.

[13] MENON Bruno. « Les langages documentaires. Un panorama, quelques remarques critiques et un essai de bilan ». *Documentaliste – Sciences de l'information*, 2007, vol. 44, n° 1, pp. 18-28.

RECHERCHE D'INFORMATION

[14] BOYER Anne. « Analyse des usages pour améliorer l'accès aux ressources ». *In* Calderan Lisette, Bernard Hidoine et Jacques Millet (coord.), *Métadonnées : mutations et perspectives*. Paris, ADBS, 2008, pp. 89-111.

[15] KOLMAYER Elisabeth. *Démarche d'interrogation documentaire et navigation*. Quatrième colloque hypermédias et apprentissages, pp. 121-134.

Références complémentaires

SIMONNOT Brigitte. « De la pertinence à l'utilité en recherche d'information : le cas du Web ». <http://hal.ccsd.cnrs.fr/docs/00/06/26/04/PDF/sic_00001410.pdf> (consulté le 23 septembre 2009).

VALORISATION DE L'INFORMATION ET WEB

[16] CALDERAN Lisette, Bernard HIDOINE et Jacques MILLET (coord.). *Métadonnées : mutations et perspectives*. Paris, ADBS, 2008

[17] CHARLET Jean, Bruno BACHIMONT, Raphaël TRONCY. « Ontologies pour le Web sémantique ». Information- Interaction- Intelligence. [en ligne]., numéro hors série « Web sémantique », 2004. <http://www.revue-i3.org/hors_serie/annee2004/revue_i3_hs2004_01_04.pdf> (consulté le 22 septembre 2009).

[18] MARCHEIX Loraine. *Conception d'une ontologie à partir d'un thésaurus spécialisé dans le domaine de l'archéologie et des sciences de l'antiquité*. Mémoire présenté pour le Master II professionnel de Gestion de l'information et du document, spécialité gestion des connaissances, Saint-Denis, Université Vincennes-Saint-Denis, Paris 8, 2009.

[19] MENON Bruno. « Journée d'étude ADBS – IUT Paris V, Le web sémantique : de nouveaux enjeux documentaires ? ». *Documentaliste-Sciences de l'information*, 2003, vol. 40, n° 6, pp. 387-391.

[20] MENON Bruno. *Technologies de l'information. Web sémantique*. INTD-CNAM Juin 2008

[21] MESGUISCH Véronique. « Le Web sémantique : utopie ou réalité ? ». *Netsources*, Novembre/ Décembre, 2007, n° 71, pp. 1-5.

Références complémentaires

Documentaliste – Sciences de l'information. Dossier Web 2.0 & information-documentation. 2009, vol. 46, n° 1.

EDITION PAPIER/EDITION ELECTRONIQUE

Références complémentaires

GRESILLAUD Sylvie, Jacqueline GILLET et Catherine MOREL-PAIR. « L'édition électronique scientifique, son impact dans le traitement de l'information d'un centre de documentation ». *In* Broudoux Évelyne et Ghislaine Chartron (dir.). *In* Broudoux Évelyne et Ghislaine Chartron (dir.). *Traitements et pratique documentaires : vers un changement de paradigme ? Actes de la deuxième conférence Document numérique et Société, Paris, CNAM, 17-18 novembre 2008*. Paris, ADBS éditions, pp. 279-297.

IBEKWE-SAN JUAN Fidelia. « Repérage et annotation d'indices de nouveautés dans les écrits scientifiques ». *In* Timimi Ismaïl et Susan Kovacs (coord.). *Indice, index, indexation. Actes du colloque international organisé les 3 et 4 novembre 2005 à l'Université de Lille-3 par les laboratoires Cersates et Gerico*. Paris/Lille, ADBS éditions/Université de Charles-de-Gaulle Lille-3/Gérico, 2006, pp. 261-275.

PROST Bernard. *Rapport d'étude sur l'édition numérique de livres scientifiques et techniques. L'éditeur des années 2010*. Paris, Etude réalisée par QUAE avec le soutien du Ministère de l'enseignement supérieur et de la recherche, 2007, 98 p.

RIGOT Huguette. « Ordre/désordre de l'écrit : le chercheur en sciences sociales et humaines, entre imprimé et numérique ». *In* Timimi Ismaïl et Susan Kovacs (coord.). *Indice, index, indexation. Actes du colloque international organisé les 3 et 4 novembre 2005 à l'Université de Lille-3 par les laboratoires Cersates et Gerico*. Paris/Lille, ADBS éditions/Université de Charles-de-Gaulle Lille-3/Gérico, 2006, pp. 277-291.

SMOLCZEWSKA-TONA Agnieszka et Geneviève LALLICH-BOIDIN. « De l'édition tradictionnelle à l'éditions numérique : le cas d la presse du XIX^{e} siècle ». *In* Broudoux Évelyne et Ghislaine Chartron (dir.). *Traitements et pratique documentaires : vers un changement de paradigme ? Actes de la deuxième conférence Document numérique et Société, Paris, CNAM, 17-18 novembre 2008*. Paris, ADBS éditions, pp. 299-316.

Annexes

Annexe 1 – Liste des abréviations

classées par ordre alphabétique

AES	*Archives européennes de sociologie*
BAT	Bon à tirer
DSI	Diffusion sélective d'information
CNRS	Centre national de la recherche scientifique
CUP	Cambridge University Press
JO	*Journal officiel*
JSA	*Journal de la Société des Américanistes*
LESC	Laboratoire d'ethnologie et de sociologie comparative
MAE	Maison Archéologie et Ethnologie ou Maison René-Ginouvès, Nanterre
SHS	Sciences Humaines et Sociales
SR	Secrétaire de rédaction
RA	*Revue archéologique*

Annexe 2 – Extrait du Rapport d'activités 2007 de la Maison René-Ginouvès

En confiant la « *gestion* » de deux **revues** propres du CNRS, les revues *Gallia* et *Paléorient*, à l'UMS 844 dès sa création, les responsables du CNRS nous ont chargés d'une mission qui entraîne pour nous une lourde responsabilité. Et ce d'autant plus que dans le même temps, et notamment au cours du dernier quadriennal, deux revues, *Archives Européennes de Sociologie (AES)* et *Ethnologie Française*, sont venues rejoindre le *Journal de la Société des Américanistes (JSA)*, pour constituer un groupe de revues aidées par le CNRS. Pour ce dernier ensemble, les moyens mis en œuvre par le CNRS sont humains dans les trois cas et également financiers pour le *JSA*. À l'heure où nous rédigeons ce projet de quadriennal, nous pouvons dire que d'autres revues sont susceptibles de se localiser très prochainement à la Maison René-Ginouvès.

Dans le même temps du quadriennal 2005-2008, la Maison a pris une initiative éditoriale, liée à la tenue de nos colloques annuels dont nous éditons chaque année les actes. Ainsi ont été publiés les **« Colloques de la Maison René-Ginouvès »**. A la liste proposée ci-dessus, nous pouvons annoncer la préparation des actes de notre colloque 2007, « L'Eau : enjeux, usages et représentations ». Ces actes sont publiés dans les onze mois qui suivent la rencontre, soit juste avant la tenue du colloque annuel. Ceci est possible car la secrétaire de rédaction qui travaille pour les *AES* est employée par cette revue à mi-temps, l'autre mi-temps étant affecté à cette édition.

Parallèlement nous avons lancé la **collection « Travaux de la Maison René-Ginouvès »** qui édite des monographies ou des travaux collectifs (liste ci-dessus). Un Comité de lecture est en train d'être mis en place pour le choix des prochains ouvrages. Le projet, déjà accepté par notre Comité de direction et qui sera soumis au Conseil scientifique du 16 novembre 2007, consiste à confier la mission de comité de lecture au Conseil scientifique, parmi lequel figurent trois étrangers ainsi que trois membres de la MAE et trois Français hors-MAE. Celui-ci pourra donc émettre ses avis à partir des rapports de deux experts lecteurs extérieurs. Pour ne pas retarder les publications, il pourrait fonctionner à distance, et non seulement chaque année lors de la réunion du Conseil scientifique.

Chacune des revues a son équipe éditoriale, constituée de six personnes à *Gallia*, et d'une personne pour chacune des autres revues (un mi-temps pour les *AES*). Animés du souci de donner cohérence à cet ensemble, nous réfléchissons à une forme de structuration dans ce

que nous pourrions reconnaître comme un véritable **« Pôle éditorial »**. Ce projet est nouveau et sera discuté avec tous les partenaires de la Maison, au Comité de direction et avec notre tutelle. Notre tutelle dans ce domaine, ici le seul CNRS, est confrontée à une difficulté de personnel. Avec le Directeur scientifique adjoint en charge de la Maison René-Ginouvès, Monsieur Bruno Laurioux, nous nous sommes efforcés, notamment au cours d'une réunion tenue le 27 juin 2007, de décliner les différentes étapes et activités dans la fabrication d'une revue afin d'identifier la ou les activités qui sont susceptibles d'être mutualisées, du choix de l'article à la diffusion. Au cours de cette réunion, et après mûre réflexion, il est apparu que deux tâches pourraient prioritairement être mutualisées : mise en page et infographie ; la mise en ligne pouvant l'être aussi. Et c'est donc dans cet esprit que nous formulerons la première demande de recrutement pour le Pôle éditorial.

Annexe 3 – Pôle documentaire

Bibliothèque d'archéologie

Présentation	Spécialisée en préhistoire, archéologie et sciences du monde antique et médiéval, la bibliothèque est née en 1997 du rassemblement de fonds documentaires constitués par plusieurs équipes du Centre national de la recherche scientifique (CNRS), des Universités de Paris 1 et de Paris X. La bibliothèque participe au réseau national FRANTIQ et au réseau régional Archéologie Ile-de-France Des fonds importants ont enrichi successivement ces collections.
Domaines couverts	Préhistoire et Protohistoire européenne ; Protohistoire égéenne ; Antiquité gréco-romaine ; Moyen Age ; Littérature et philologie grecques et latines ; Sources chrétiennes ; Archéologie du Proche et Moyen Orient ; Assyriologie ; Archéologie d'Asie Centrale ; Archéologie d'Afrique orientale ; Sciences et technologies appliquées à l'archéologie ; Archéologie environnementale ; Archéometrie
Fonds	35 000 monographies ; 1 300 périodiques dont 400 vivants ; 11 000 tirés à part en accès indirect
Accès	http://www.mae.u-paris10.fr/siteums/menubiblio.php?ID=8= Du lundi à jeudi de 9 h 30 à 18h, Vendredi de 9 h 30 à 17 h 30. Pas de fermeture annuelle. La bibliothèque est ouverte aux chercheurs et enseignants du CNRS et des Universités, personnels techniques et administratifs de la MAE, étudiants après la licence ; Documents en libre accès à l'exception des tirés à part et de quelques périodiques. Consultation sur place uniquement (pas de prêt à domicile) et consultation du catalogue en ligne, des ressources électroniques de l'Université Paris-X. Accès aux catalogues collectifs des réseaux FRANTIQ et Archéologie-Ile-de-France. Prêt entre bibliothèques 45 places assises avec prises Internet ou wi-fi 5 postes de consultation : catalogue, bases de données, signets 1 photocopieur
Outils / Ressources en ligne	-Base de données sur KOHA **Ressources électroniques disponibles dans la salle de lecture :** -Accès aux bases de données et périodiques en ligne du réseau de l'Université Paris X-Nanterre -Accès à Biblio-SHS : portail développé par l'INIST – JSTOR, FRANCIS, Arts & Humanities Citation index, Bulletin Analytique d'Histoire Romaine, Dissertation Abstracts, Historical Abstracts, DRANT (DRoits ANTiques), Social Sciences Citation Index, etc., Dyabola, Bibliotheca Teubneriana Latina, Thesaurus Linguae Graecae (TLG), Archives Bankes (Dorset). -Accès aux bases de données et aux publications produites à la Maison René Ginouvès, et aux sites du Portail des Maisons des Sciences de l'Homme **Bases et revues produites à la Maison :** Cahiers des thèmes transversaux ArScAn : Résumés des séminaires organisés par les thèmes de l'UMR ; La Chronique Internet des Nouvelles de l'archéologie ; Gallia, Gallia Préhistoire et leurs suppléments, Gallia Informations ; LIMC-France (Lexicon Iconographicum Mythologiae Classicae) ; Paléorient : résumés français/anglais des articles parus et bibliographie annuelle
Equipe	2 responsables + 1 magasinier + 2 bibliothécaires + 2 vacataires

Bibliothèque Éric de Dampierre, du laboratoire d'ethnologie et de sociologie comparative

Présentation	Créée en 1967, sous l'impulsion du Professeur de Dampierre, fondateur, et alors directeur, du Laboratoire d'ethnologie et de sociologie comparative, cette bibliothèque a été conçue pour soutenir les programmes scientifiques et les enseignements appelés à se développer au sein des formations d'ethnologie, de préhistoire et de sociologie comparative de l'Université de Paris X. Elle figure aujourd'hui parmi les trois plus importants centres documentaires d'anthropologie français. La politique d'acquisition, arrêtée dès l'origine, privilégie les textes fondateurs des disciplines réunies ainsi que les ouvrages en langue originale. La bibliothèque a, au cours du temps, élargi sa couverture documentaire, notamment, par l'accueil de fonds spécifiques et la création d'une section des archives scientifiques Dès les années 1980, s'est constituée une base de données bibliographiques conçue avec un double niveau de descripteurs pour servir également de catalogue informatisé. Aujourd'hui ce catalogue, au format USMARC, est consultable en ligne (voir Catalogue du réseau ethnologie) Les fonds d'archives, et plus particulièrement les matériaux de terrain, sont actuellement en cours de numérisation et seront prochainement accessibles par l'intermédiaire d'une banque d'archives ethnographiques.
Domaines couverts	Ethnologie et sociologie
Fonds	d'imprimés (31 200 ouvrages, 149 collections de périodiques, dont 120 vivantes, 4 500 tirés à part) de manuscrits (12 fonds d'archives scientifiques dont huit principalement composés de matériaux de terrain)
Accès	http://www.mae.u-paris10.fr/siteums/ethnobiblio.php?ID=29= très largement ouverte à l'ensemble des chercheurs et des doctorants
Outils/ Ressources en ligne	-Base de données sur KOHA -Le catalogue des périodiques, la liste des acquisitions, les publications de chercheurs -Le réseau ethnologie est un réseau documentaire thématique et repose sur la collaboration de quatre bibliothèques : Documentation du Centre de Recherches Bretonne et Celtique, Centre de documentation André-Georges Haudricourt, Bibliothèque du Laboratoire d'anthropologie sociale, Bibliothèque du Laboratoire d'ethnologie et de sociologie comparative. Il est principalement orienté sur la production du catalogue commun, base de donnée documentaire de plus de 130 000 notices bibliographiques. -Le thésaurus GeoEthno (thésaurus géographique pour l'ethnologie)
Equipe	1 responsable et 3 bibliothécaires documentalistes

Documentation et Frantiq

Présentation	La documentation gère plusieurs projets : FRANTIQ, DAPHNÉ et « Archéologie en Ile-de-France ». Frantiq (Fédération et ressources sur l'Antiquité) est une coopérative de centres de recherches du CNRS, des universités et du Ministère de la culture, mettant en commun des bases de données sur les sciences de l'Antiquité, de la préhistoire au Moyen Âge. Le Catalogue Collectif Indexé (CCI) comprend 22 Bibliothèques productrices de données signalétiques indexées avec le thésaurus PACTOLS. En plus de ces catalogues de bibliothèque, sont intégrées les Tables des revues Gallia depuis 1983, les rapports et projets de fouilles françaises à l'étranger, et la signalétique de documents uniquement électroniques. DAPHNE est une ACI du réseau des Maison des Sciences de l'Homme sur trois années 2004-2007.
Domaines couverts	Sciences de l'Antiquité, de la préhistoire au Moyen Age
Accès	En ligne : http://www.mae.u-paris10.fr/siteums/menufrantiq.php?ID=24=
Outils/ Ressources en ligne	-Frantiq est passé sous koha en catalogage partagé -DAPHNE (Données en Archéologie, Préhistoire et Histoire sur le NEt) est un guichet unique d'accès gratuit à des bases de données bibliographiques thématiques. Thèmes : préhistoire, protohistoire, archéologie (depuis les premiers témoignages humains jusqu'à l'ère industrielle), sciences de l'Antiquité (sous tous ses aspects : religion, politique, société, philosophie, économie, art, vie quotidienne, etc.) et histoire jusqu'à l'an Mil, sur tous les continents. Nature des données : notices bibliographiques complémentaires au niveau des types de documents analysés (articles, monographies, actes de congrès, expositions, thèses, revues, ressources électroniques…) à partir de la littérature internationale et de la finesse des niveaux d'analyse (indexation, voire résumé). Spécificités : L'idée centrale sur laquelle repose ce portail est de fournir des réponses pertinentes à des questions ciblées grâce à l'utilisation des langages documentaires précis et actualisés. Ainsi Daphné est une mutualisation de données, c'est-à-dire une convergence de trois bases de données BAHR, FRANCIS (domaines Pré- et protohistoire, art et archéologie et partiellement histoire et sciences des religions), FRANTIQ-CCI.
Equipe	2 documentalistes

Service des archives

Présentation	Crée en 1998 en même temps que la Maison René-Ginouvès, ce service a pour mission la collecte d'archives historiques. À ce jour, le service conserve 207,60 mètres linéaires d'archives déposées par les unités, les équipes de recherche et les chercheurs de la Maison René-Ginouvès. Les archives traitées représentent 161,20 mètres linéaires.
Domaines couverts	archives administratives et scientifiques d'unités de recherche d'archéologie, d'équipes et de chercheurs
Fonds	Plusieurs fonds parmi lesquels on trouve celui de Jacques Tixier, 1952-1989 ; Joëlle Burnouf, 1953-2000 ; Équipe Archéologie de l'Asie centrale et Jean-Claude Gardin, 1955-1996 ; Documentation en archéologie-Brigitte Lequeux, 1967-1999 ; Alain Marliac, 1968-2001 ; Odette Sargnon-Bijoux préhelléniques, ant.1970 ; Équipe Protohistoire européenne, 1970-2004 ; Mission préhistorique française au Kenya-Hélène Roche, 1975-2003 ; Laboratoire Préhistoire et Technologie, 1977-2004 ; Luc Bachelot, 1984-2002
Accès	Base de données diffusée sur le Web ; site web en cours de fabrication
Outils/ Ressources en ligne	Classement, inventaire, tableaux de gestion, bordereau de versement… Arkeïa couplé avec pléade
Equipe	1 archiviste et 1vacataire

Annexe 4 – Mots-clés / descripteurs

Dans le tableau suivant, vont être présentés quelques exemples de mots-clés géographiques qui ont été pris dans le JSA et qui vont être rapprochés des descripteurs du thésaurus Géoethno.

Index JSA Géographiques / ethniques	Occurrences dans le JSA	Dans le JSA : précisions apportées dans		Dans Geoethno : précisions fournies ou distinctions faites	Correspondance thésaurus/index
		le Titre, résumé et/ou texte	les Mots-clés du texte		
Amazonie	13 dont on retiendra 6 exemples	**Texte 1–**Nord-Ouest amazonien (titre)	**Texte 1–**Amazonie du Nord-Ouest	8 réponses : 1. Amazonie, FORET, //Trans-Amérique 2. Amazonie colombienne, REGION Amérique/Amérique du Sud/ Colombie 3. Brésil, Pays ou territoire/ Amérique/Amérique du Sud 4. Colombie, Pays ou territoire ancien(s) nom(s) : Nouvelle-Grenade/Amérique/Amérique du Sud 5. Guyane française, Pays ou territoire/Amérique/Amérique du Sud 6. Haute Amazonie, REGION //Trans-Amérique 7. Pérou, Pays ou territoire/ Amérique/Amérique du Sud 8. République de l'Equateur, Pays ou territoire/Amérique/Amérique du Sud	**Texte 1–** Difficile à définir : il aurait fallu demander à l'auteur
		Texte 2–Amazonie brésilienne	**Texte 2–**Amazonie Indexé aussi à Brésil et Pérou		**Texte 2–** 3. Brésil, Pays ou territoire/ Amérique/Amérique du Sud 7. Pérou, Pays ou territoire/ Amérique/Amérique du Sud réindexation possible : Amazonie, Brésil, Pérou
		Texte 3–juste le groupe ethnique précisé : tupi-guarani	**Texte 3–**		**Texte 3–** Difficile à définir : il aurait fallu demander à l'auteur
		Texte 4–Amazonie méridionale	**Texte 4–**indexé aussi à Brésil		**Texte 4–** 3. Brésil, Pays ou territoire/ Amérique/Amérique du Sud réindexation possible : Amazonie, Brésil
		Texte 5– Amazonie occidentale (titre), rives péruviennes du fleuve purus, au cœur du bassin amazonien (texte)	**Texte 5–**	Le fleuve Purus pas de réponse dans Géoethno	**Texte 5–** 7. Pérou, Pays ou territoire/ Amérique/Amérique du Sud réindexation possible : Amazonie, Brésil, fleuve Purus
		Texte 6–Amazonie (titre), Brésil central (résumé), dans l'interfluve Xingu-Tocantins, dans l'État du Pará, au Brésil (texte)	**Texte 6–**	1 entrée à Pará Type : Zone administrative ; Employé pour BR-PA ; Terme(s) générique(s) : Nord – BRA ; Niveau générique : Amérique du Sud/Brésil/Nord – BRA ; Terme(s) spécifique(s) : Belém (URB), Marajó (ILE),Obidos (URB), Santarém (URB) Xingu-Tocantins pas de réponse dans Géoethno	**Texte 6–** 3. Brésil, Pays ou territoire/ Amérique/Amérique du Sud réindexation possible : Amazonie, Brésil, Pará

Index JSA Géographiques / ethniques	Occurrences dans le JSA	Dans le JSA : précisions apportées dans le Titre, résumé et/ou texte	Dans le JSA : précisions apportées dans les Mots-clés du texte	Dans Geoethno : précisions fournies ou distinctions faites	Correspondance thésaurus/index
Antilles	1	les Petites antilles	antilles précolombiennes	29 réponses	22. Petites Antilles, ARCHIPEL/Océan Atlantique/Mer des Caraïbes
Ausangate	1	cordillère de l'Ausangate (Andes sud-péruviennes) (résumé)	Andes, Pérou	0 Ausangate pas de réponse dans Géoethno	
Baie James	1	Est de la Baie James (titre), Amérique du nord (texte)		2 Réponses 1. Baie James, BAIE, Amérique du Nord/Canada/Baie d'Hudson 2. Nord du Québec, REGION, Amérique du Nord/Canada/Province du Québec	1. Baie James, BAIE, Amérique du Nord/Canada/Baie d'Hudson
Campeche	4 dont on retiendra 2 exemples	**Texte 1–**Rio Bec, Katjún=nom de site (titre) ; État de Campeche (texte)	**Texte 1–**	4 Réponses à Campeche 1. Campeche, Ville ou village, Péninsule du Yucatán/État de Campeche/Municipe de Campeche 2. État de Campeche, Zone administrative, Amérique centrale/Mexique/Péninsule du Yucatán 3. Golfe de Campeche, GOLFE, Amérique/Amérique centrale/Mexique 4. Municipe de Campeche, Zone administrative, Mexique/Péninsule du Yucatán/État de Campeche Kajtún pas de réponse dans Géoethno	**Texte 1–** 2. État de Campeche, Zone administrative, Amérique centrale/Mexique/Péninsule du Yucatán Rio bec voir plus bas Mésoamérique voir plus bas réindexation possible : Mésoamérique, Mexique, État de Campeche, Rio bec, Kajtún
		Texte 2–Rio Bec (texte)	**Texte 2–**		**Texte 2–** 2. État de Campeche, Zone administrative, Amérique centrale/Mexique/Péninsule du Yucatán Rio bec voir plus bas Mésoamérique voir plus bas réindexation possible : Mésoamérique, Mexique, État de Campeche, Rio bec

Index JSA Géographiques / ethniques	Occurrences dans le JSA	Dans le JSA : précisions apportées dans le Titre, résumé et/ou texte	Dans le JSA : précisions apportées dans les Mots-clés du texte	Dans Geoethno : précisions fournies ou distinctions faites	Correspondance thésaurus/index
Chaco	2	**Texte 1–**Chaco argentin et bolivien	**Texte 1–**	7 Réponses 1. Argentine, Pays ou territoire, ancien(s) nom(s) : Provinces Unies du Rio de la Plata/Amérique/Amérique du Sud 2. Bolivie, Pays ou territoire, /Amérique/Amérique du Sud 3. Chaco, Zone administrative, Amérique/Amérique du Sud/Argentine 4. Chaco Canyon, Site archéologique, culturel ou naturel, Amérique du Nord/Etats-Unis/Nouveau-Mexique 5. Gran Chaco, REGION, //Trans-Amérique 6. Paraguay, Pays ou territoire, /Amérique/Amérique du Sud 7. Salta, Zone administrative, Amérique/Amérique du Sud/Argentine Détail des réponses pour le Chaco Type Zone administrative Employé pour AR-K Terrain de chasse Terme(s) générique(s) Argentine Niveau générique : Amérique/Amérique du Sud/Argentine	**Texte 1–** réindexation possible : Chaco, Bolivie, Argentine
		Texte 2–Bolivie (résumé)	**Texte 2–**		**Texte 2–** réindexation possible : Chaco, Bolivie

Index JSA Géographiques / ethniques	Occurrences dans le JSA	Dans le JSA : précisions apportées dans		Dans Geoethno : précisions fournies ou distinctions faites	Correspondance thésaurus/index
		le Titre, résumé et/ou texte	les Mots-clés du texte		
Gran Nayar	2	**Texte 1–**Ouest du Mexique (résumé)	**Texte 1–**Mexique	0 Gran Nayar pas de réponse dans Géoethno En revanche, on trouve Nayarit qui est une État du Mexique, mais qui ne couvre pas exactement la même zone que le Gran Nayar Détail des réponses pour Mexique Ancien(s) nom(s) : Nouvelle-Espagne Type : Pays ou territoire Employé pour Estados Unidos Mexicanos, Etats-Unis du Mexique, MEX, MX, Mexico, Mexiko Méjico, México [...] Zone(s) transfrontalière(s) : Aire Maya, Basses terres Maya, Caraïbe, Golfe du Mexique, Hautes terres Maya, Mésoamérique, Oasisamérique, Océan Atlantique, Océan Pacifique, Rio Grande, Rio Usumacinta, Sierra Madre Terme(s) générique(s) : Amérique centrale Niveau générique : /Amérique/Amérique centrale Terme(s) spécifique(s) : Aguascalientes (ADM), Aridamérique (REGION), Baja California (ADM), Baja California Sur (ADM), Chiapas (ADM), Chihuahua (ADM), Coahuila (ADM), Colima (ADM), Durango (ADM), Golfe de Californie (GOLFE), Golfe de Campeche (GOLFE) [...], Jalisco (ADM), Mexique précolombien (HISTOIRE) (ADM), Michoacán (ADM), Morelos (ADM), Nayarit (ADM), [...] San Luis Potosí (ADM), Sierra Madre du Sud (MONT), Sierra Madre occidentale (MONT), Sierra Madre orientale (MONT) .../...	**Texte 1–** Difficile à définir : il aurait fallu demander à l'auteur
		Texte 2–Nord du Mexique (titre) ; ouest et nord-ouest du Mexique (résumé)	**Texte 2–**Mexique		**Texte 2–** Difficile à définir : il aurait fallu demander à l'auteur

Index JSA Géographiques / ethniques	Occurrences dans le JSA	Dans le JSA : précisions apportées dans		Dans Geoethno : précisions fournies ou distinctions faites	Correspondance thésaurus/index
		le Titre, résumé et/ou texte	les Mots-clés du texte		
Grande Nicoya	1	nord-ouest du Costa Rica (résumé)	Costa Rica, Nicaragua	0 Grande Nicoya pas de réponse dans Géoethno Détail des réponses pour Costa Rica Type : Pays ou territoire Employé pour CR, CRI, Republic of Costa Rica, République du Costa Rica Zone(s) transfrontalière(s) : Caraïbe, Mésoamérique, Océan Atlantique, Océan Pacifique Terme(s) générique(s) : Amérique centrale Niveau générique : /Amérique/Amérique centrale Terme(s) spécifique(s) : Alajuela (ADM), Cartago (ADM), Guanacaste (ADM), Heredia (ADM), Limon (ADM), Province de San José -- CRI (ADM), Puntarenas (ADM) Détail des réponses pour Nicaragua Type : Pays ou territoire Employé pour NI, NIC, Republic of Nicaragua, République du Nicaragua Zone(s) transfrontalière(s) : Caraïbe, Mésoamérique, Océan Atlantique, Océan Pacifique Terme(s) générique(s) : Amérique centrale Niveau générique : /Amérique/Amérique centrale Terme(s) spécifique(s) : Atlántico Norte (ADM), Atlántico Sur (ADM), Boaco (ADM), Carazo (ADM), Chinandega (ADM), Chontales (ADM), Département de León (ADM), Département de Managua (ADM), Estelí (ADM), Granada (ADM), Jinotega (ADM), Madriz (ADM), Masaya (ADM), Matagalpa (ADM), Nueva Segovia (ADM), Rivas (ADM), Río San Juan (ADM)	Difficile à définir : il aurait fallu demander à l'auteur

Index JSA Géographiques / ethniques	Occurrences dans le JSA	Dans le JSA : précisions apportées dans		Dans Geoethno : précisions fournies ou distinctions faites	Correspondance thésaurus/index
		le Titre, résumé et/ou texte	les Mots-clés du texte		
Guerrero	1	Guerrero, Mexique (titre) ; municipalité tlapanèque (résumé)		Détail des réponses pour Guerrero Type : Zone administrative Employé pour (Gro.), Estado de Guerrero, MX-GRO Terme(s) générique(s) : Mexique Niveau générique : Amérique/Amérique centrale/Mexique Terme(s) spécifique(s) : Région Acapulco (ADM), Région Centro (ADM), Région Costa Chica (ADM), Région Costa Grande (ADM), Région La Montaña (ADM), Région Norte (ADM), Région Tierra Caliente (ADM), Teopantecuanitlan (SITE) Municipalité tlapanèque pas de réponse dans Géoethno	Difficile à définir : il aurait fallu demander à l'auteur
Haut Xingu	1	Haut-Xingu (titre)		Détail des réponses pour Haut Xingu Type : REGION Employé pour Alto Xingu, Upper Xingu Basin Terme(s) générique(s) : Mato Grosso Niveau générique : Brésil/Centre-Ouest -- BRA/Mato Grosso	

Index JSA Géographiques / ethniques	Occurrences dans le JSA	Dans le JSA : précisions apportées dans		Dans Geoethno : précisions fournies ou distinctions faites	Correspondance thésaurus/index
		le Titre, résumé et/ou texte	les Mots-clés du texte		
Mésoamérique	22			8 Réponses 1. Belize, Pays ou territoire, ancien(s) nom(s) : Honduras britannique/ Amérique/Amérique centrale 2. Costa Rica, Pays ou territoire/ Amérique/Amérique centrale 3. El Salvador, Pays ou territoire/ Amérique/Amérique centrale 4. Guatemala, Pays ou territoire/ Amérique/Amérique centrale 5. Honduras, Pays ou territoire/ Amérique/Amérique centrale 6. Mésoamérique, REGION//Trans-Amérique 7. Mexique, Pays ou territoire, ancien(s) nom(s) : Nouvelle-Espagne /Amérique/Amérique centrale 8. Nicaragua, Pays ou territoire/ Amérique/Amérique centrale Détail des réponses pour Mésoamérique Type REGION Employé pour Mesoamerica, Méso-Amérique Terme(s) générique(s) : Trans-Amérique Niveau générique : //Trans-Amérique	Je mets cette entrée car elle est très importante dans le domaine des études américanistes. Ce terme mésoamérique renvoie non seuleemnt à un territoire, mais aussi parfois à une époque et à une aire culturelle.

Index JSA Géographiques / ethniques	Occurrences dans le JSA	Dans le JSA : précisions apportées dans		Dans Geoethno : précisions fournies ou distinctions faites	Correspondance thésaurus/index
		le Titre, résumé et/ou texte	les Mots-clés du texte		
Puebla	1	États de Puebla, Tlaxcala, Oaxaca et Veracruz au Mexique (résumé)	Cholula	4 réponses 1. État de Puebla, Zone administrative Amérique/Amérique centrale/Mexique 2. Municipe Heroica Puebla de Zaragoza, Zone administrative, Mexique/État de Puebla/Région Puebla 3. Puebla, Ville ou village, État de Puebla/Région Puebla/Municipe Heroica Puebla de Zaragoza 4. Région Puebla, Zone administrative, Amérique centrale/Mexique/État de Puebla Détail des réponses pour État de Puebla Type : Zone administrative ; Employé pour (Pue.), Estado de Puebla MX-PUE, Puebla ; Terme(s) générique(s) : Mexique Niveau générique : Amérique/Amérique centrale/Mexique ; Terme(s) spécifique(s) : Cerro Toluquilla (VOLCAN) [...], Région Puebla (ADM), Région San Pedro Cholula (ADM) .../... Détail des réponses pour Cholula Type : Site archéologique, culturel ou naturel Terme(s) générique(s) : Région San Pedro Cholula Niveau générique : Mexique/État de Puebla/Région San Pedro Cholula Détail des réponses pour Tlaxcala Type : Zone administrative Employé pour (Tlax.), Estado de Tlaxcala, MX-TLA Terme(s) générique(s) : Mexique Niveau générique : Amérique/Amérique centrale/Mexique Terme(s) spécifique(s) : Municipe Acuamanala de Miguel Hidalgo (ADM) .../...	Les termes concordent

Index JSA Géographiques / ethniques	Occurrences dans le JSA	Dans le JSA : précisions apportées dans		Dans Geoethno : précisions fournies ou distinctions faites	Correspondance thésaurus/index
		le Titre, résumé et/ou texte	les Mots-clés du texte		
Rio Bec	3	Les trois textes sont identiques aux trois premiers signalés pour l'entrée Campeche		Région du Río Bec, Type, Site archéologique, culturel ou naturel Employé pour Región Río Bec, Rio Beque, Région du Rio Bec, Zona Arqueológica Río Bec ; Terme(s) générique(s) : Municipe de Calakmul ; Niveau générique : Péninsule du Yucatán/État de Campeche/Municipe de Calakmul	
Vaupès	1	Amazonie du Nord-Ouest (résumé)		4. réponses 1. Brésil, Pays ou territoire /Amérique/Amérique du Sud 2. Colombie, Pays ou territoire, ancien(s) nom(s) : Nouvelle-Grenade/ Amérique/Amérique du Sud 3. Département du Vaupés, Zone administrative, Amérique/Amérique du Sud/Colombie 4. Río Vaupés, RIVIERE, Trans-Amérique/Amazone/Rio Negro Détail des réponses pour Département du Vaupés Type : Zone administrative ; Employé pour CO-VAU, Departamento de Vaupés, Vaupes Department, Vaupés – adm ; Terme(s) générique(s) : Colombie ; Niveau générique : Amérique/Amérique du Sud/Colombie Terme(s) spécifique(s) : Puerto Carreno (URB) Détail des réponses pour Río Vaupés Type : RIVIERE ; Employé pour Rio Caiari, Rio Caiary, Rio Cayari, Rio Uaapes, Rio Uaapés, Rio Uapes, Rio Vaupes, Rio Vaupés, Río Uaupés, Uaupés ; Terme(s) générique(s) : Rio Negro ; Niveau générique : Trans-Amérique/Amazone/Rio Negro ; Terme(s) spécifique(s) : Cuquiari (RIVIERE), Papuri (RIVIERE), Querari (RIVIERE), Tiquié (RIVIERE)	Difficile à préciser : il aurait fallu demander à l'auteur

Index JSA Géographiques / ethniques	Occurrences dans le JSA	Dans le JSA : précisions apportées dans		Dans Geoethno : précisions fournies ou distinctions faites	Correspondance thésaurus/index
		le Titre, résumé et/ou texte	les Mots-clés du texte		
Yucatan Indexé également aux entrées suivantes : Mexique, Mésoamérique, Chiapas	1	Mayas des Basses Terres (titre)		2 Réponses 1. Péninsule du Yucatán, PENINSULE, Amérique/Amérique centrale/Mexique 2. Yucatán, Zone administrative, Amérique centrale/Mexique/Péninsule du Yucatán Détail des réponses pour Péninsule du Yucatán Type : PENINSULE Employé pour Basses terres maya du nord, Northern Lowland Maya, Péninsule mexicaine, Tierras Bajas del Norte, Tierras bajas septentrionales.../... ; Terme(s) générique(s) : Mexique ; Niveau générique : Amérique/Amérique centrale/Mexique ; Terme(s) spécifique(s) : Quintana Roo (ADM) ; Yucatán (ADM) ; État de Campeche (ADM) Détail des réponses pour Yucatán Type : Zone administrative ; Employé pour (Yuc.), Estado de Yucatán, MX-YUC, Yucatán – adm ; Terme(s) générique(s) : Péninsule du Yucatán ; Niveau générique : Amérique centrale/Mexique/Péninsule du Yucatán ; Terme(s) spécifique(s) : Acancéh (SITE), Aké (SITE), Balancanché (SITE), Chacmultún (SITE), Chan Kom (URB), Chichén Itzá (SITE), Cratère de Chicxulub (SITE), Dzibilchaltún (SITE).../...	

Annexe 5 – Questionnaires

QUESTIONNAIRE POUR LES AUTEURS

Ce questionnaire va se faire en deux temps : Le premier vous placera en tant que chercheur, enseignant-chercheur, étudiant ayant besoin de chercher de l'information sur un sujet dans le but d'écrire un texte ou de préparer un cours ou autre ? Le second vous placera en tant qu'auteur d'un texte.

RECHERCHE D'INFORMATIONS

Procédés mis en place pour une recherche d'information

Comment prenez-vous connaissance d'un texte intéressant pour vous qui vient d'être publié dans une revue ?

Pouvez-vous définir les cas de figure qui vous conduisent à rechercher des informations en général, puis plus particulièrement dans une revue ?

Pouvez-vous définir le type d'information que vous pouvez être amené à chercher en général, puis plus particulièrement dans une revue ?

En fonction de ce type d'informations (texte, références bibliographiques, illustrations…) comment procédez-vous pour les chercher en général, puis plus particulièrement dans une revue papier ? et dans sa version électronique ?

Quel type d'outils (informatiques ou non) avez-vous l'habitude d'utiliser pour ce type de recherche ?

Pratiquez-vous beaucoup les outils informatiques ? Si oui, lesquels

Connaissez-vous des bases de données pouvant vous aider à faire ces recherches ? Lesquelles ?

À quel rythme avez-vous l'habitude de les interroger ?

Utilisez-vous d'autres outils papiers et/ou électroniques pour ces recherches ? si oui lesquels et comment les utilisez-vous ?

Connaissez-vous des thésaurus ? si oui lesquels ? selon vous, à quoi servent-ils ?

Selon vous, à quoi servent les mots-clés ? ET/OU à quoi devraient-ils servir ?

Selon vous, à quoi servent les index ? les utilisez-vous ?

Résultats de la recherche

Êtes-vous satisfait des résultats obtenus en fontion de vos modes de recherche ?
Toujours souvent jamais

Pouvez-vous expliquer en quelques mots ce qui vous paraît bien, moins bien et très mal fonctionner dans vos recherches telles que vous les pratiquez ?

AUTEUR DE TEXTE

À quel moment de la rédaction de votre texte avez-vous commencé à penser aux mots-clés ?

Comment avez-vous procédé pour construire vos mots-clés en fonction du texte rédigé ? quelles questions vous posez-vous pour les construire ?

Selon vous, à quoi doivent-ils servir ?

Êtes-vous doctorant ?
chercheurs (chercheurs, enseignants-chercheurs, chercheurs hors statut) ?
professionnels de l'information (documentalistes, bibliothécaires, journalistes) ?
professionnels de l'édition (secrétaire de rédaction, éditeur, rédacteur, directeur) ?

QUESTIONNAIRE POUR LES SECRETAIRES DE REDACTION, DIRECTEURS ET/OU REDACTEURS DE REVUE

L'axe que j'ai choisi d'aborder au cours de ces entretiens est celui de la valorisation des revues. Par valorisation j'entends tout ce qui permet de rendre plus visible et plus accessible le contenu d'une revue. Il s'agit d'essayer de déterminer les besoins en matière de valorisation pour chaque revue, puis pour l'ensemble des revues du pôle.

Selon vous, qu'est-ce qui dans vos pratiques éditoriales va permettre de valoriser la revue pour laquelle vous travaillez ?

Utilisez-vous des bases de données pour votre revue ? en avez-vous une ou plusieurs ? Pour quelle type de données ? Sont-elles accessibles sur le net ? dans quel but ?

Utilisez-vous des tables ? en avez-vous ? De quelle façon ?

Utilisez-vous des index ? en avez-vous ?

Avez-vous déjà un site web ? pouvez-vous en faire une description sur ce qu'il y a, ce qui manque, ce qui pourrait être amélioré ?

Utilisez-vous d'autres outils qui vont dans ce sens ? Lesquels ? À quel moment les utilisez-vous ? Comment les utilisez-vous ?

Avez-vous d'autres idées qui iraient dans le sens d'une valorisation de la revue ?

Demandez-vous des mots-clés à vos auteurs ? combien ? en quelle langue ? pour quel type de textes (article, compte rendu ou autre selon les rubriques de la revue) ? les vérifiez-vous ? les corrigez-vous ? en ajoutez-vous ? en proposez-vous d'autres ? si oui, pourquoi et comment, en fonction de quels critères ?

Selon vous, à quoi servent-ils ? à quoi devraient-ils servir ?

Pensez-vous qu'il serait possible de mettre en place une politique de valorisation de l'ensemble des revues du pôle ? si oui, comment l'envisagez-vous ?