



HAL
open science

Internet, nouveaux usages de la recherche dans le domaine hispanique

Marc Marti

► **To cite this version:**

Marc Marti. Internet, nouveaux usages de la recherche dans le domaine hispanique. *Les Langues néo-latines : revue de langues vivantes romanes*, 2007, 340, pp.63-78. halshs-00605064

HAL Id: halshs-00605064

<https://shs.hal.science/halshs-00605064>

Submitted on 30 Jun 2011

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

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

Internet : nouveaux usages de la recherche dans le domaine hispanique

L'usage de l'Internet a modifié les pratiques de la recherche au cours des dix dernières années. Le changement se manifeste de plusieurs façons. Dans les lignes qui suivent, nous proposons d'analyser les nouvelles modalités d'accès aux documents, les nouveaux gestes qu'elles induisent pour le chercheur dans le domaine hispanique, les limitations et les perspectives du développement d'un Internet scientifique. La réflexion sera menée à partir du domaine péninsulaire, qui est celui que nous connaissons le mieux.

L'accès aux documents a été bouleversé par le passage aux supports numériques, qui s'est traduit par l'émergence de phénomènes en partie contradictoires. En effet, la numérisation combinée avec l'accès Internet engendre une croissante accessibilité des documents qui facilite le travail de recherche. Cependant, la hausse constante de la masse documentaire disponible vient agir négativement, noyant en partie l'information pertinente dans une structure peu hiérarchisée (Internet) et inorganisée par essence.

Ce phénomène est parfois amplifié par le manque de formation à la recherche documentaire électronique, souvent pratiquée en autodidacte : une des grandes interrogations est comment utiliser Internet de manière efficace ? Il conviendra d'ébaucher un schéma qui prenne en compte à la fois les outils existants et la référence indispensable à la recherche « papier » qui est loin d'être supplantée. Les spécificités des ressources de l'hispanisme devront aussi être considérées. En effet, l'action institutionnelle et privée a produit des résultats différents dans la péninsule et en France.

Finalement, il conviendra de s'interroger sur l'état actuel des outils et des méthodes de la recherche par Internet dans le domaine hispanique, en insistant, provisoirement, sur les limites du « tout en ligne » et sur la mutation en cours. Une réflexion qui démontre que les lignes qui suivent pourraient être rapidement frappées d'obsolescence, le mal qui guette tout ce qui touche de près ou de loin aux technologies de l'information et de la communication...

1. Le passage au numérique : un accès facilité à la documentation ?

1.1. Les bibliothèques virtuelles, France, Espagne et projets internationaux

Le monde numérique s'est réellement développé à la fin des années quatre-vingts. D'une part, l'arrivée de supports comme le cédérom permettait de stocker une masse importante de documents numérisés, d'autre part, le développement et la mise en réseau des catalogues de bibliothèques augmentaient le nombre de références documentaires disponibles. L'essor d'Internet, qui facilitait une très large diffusion de l'ensemble, a accéléré le mouvement.

Les projets fondateurs ont souvent été orientés vers la diffusion de documents de référence et la création d'une bibliothèque d'utilité publique qui contiendrait les textes majeurs d'une culture (mondiale ou nationale). Le projet Gutenberg est exemplaire sur ce point. Lancé en 1971, « il se concentre sur une littérature historiquement significative et les œuvres de référence »¹. Le slogan du projet « brisons les barrières de l'ignorance et de l'illettrisme » résume le propos altruiste des fondateurs. Fonctionnant sur la base du volontariat, il réunit actuellement environ 16 000 textes en plusieurs langues.

Les initiatives institutionnelles ont oscillé entre le désir de créer une bibliothèque patrimoniale de « classiques » et la nécessité de rendre disponible des documents d'accès difficile par leur rareté ou leur fragilité. Ces deux tendances peuvent sembler contradictoires : les classiques sont par essence les livres les mieux diffusés et ils constituent le fonds de toutes les bibliothèques. Du point de vue du chercheur, ils se situent aux antipodes du document « rare » ou difficile à manipuler. En France, le fonds Gallica de la BNF reflète assez bien cette tendance : les 70 000 ouvrages proposés à la consultation en ligne « ont été choisis de manière à dessiner une bibliothèque patrimoniale et encyclopédique² ». Les autres initiatives de bibliothèque virtuelle francophones sont peu nombreuses et seules trois sont à mentionner : ABU³, ATHENA⁴ et la BIBLIOTHÈQUE ÉLECTRONIQUE DE LISIEUX⁵.

¹Voir l'article de l'encyclopédie Wikipédia à ce propos : http://fr.wikipedia.org/wiki/Projet_Gutenberg

²Présentation du fonds GALLICA <http://gallica.bnf.fr/>

³<http://abu.cnam.fr/> Il s'agit d'une bibliothèque d'auteurs francophones proposée par la CNAM. On trouve aussi des dictionnaires. Le fonds est constitué de 288 textes. Le projet date de

En Espagne, les initiatives ont été moins centralisées, générant une forte expansion des documents accessibles mais aussi une dispersion plus grande de leurs points d'accès. En effet, les communautés autonomes ont été partie prenante dans les projets de mise en ligne du patrimoine bibliographique. On retrouve donc, sans que la liste soit exhaustive, des bibliothèques virtuelles en Andalousie, au Pays Basque, à Valence, aux Canaries comptant toutes plusieurs milliers d'ouvrages à côté du fonds très modeste de la BNE (Biblioteca Nacional de España) soit 300 livres pour un total de 9 000 documents. Ici encore, la finalité patrimoniale, cette fois à l'échelon régional, a été la principale motivation.

Dans la péninsule, le projet qui, au bout du compte, est devenu la référence en la matière est issu à l'origine d'un partenariat entre l'université d'Alicante et des fondations privées : il s'agit de la *Biblioteca Virtual Miguel de Cervantes* (BVC). Les motivations semblent similaires à celles de Gallica :

« [...]es un amplio proyecto de edición digital del patrimonio bibliográfico, documental y crítico español e hispanoamericano, que pretende potenciar la expansión universal de las culturas hispánicas a través de la utilización y aplicación de los medios tecnológicos más avanzados⁶ ».

On remarque cependant à l'usage qu'une part assez importante est réservée aux études critiques (thèses, articles, collections complètes de revues scientifiques). Actuellement, la bibliothèque affiche un fonds de 14 000 titres, ce qui ne correspond pas forcément au même nombre d'ouvrages, puisqu'une revue est considérée comme un seul titre pour des dizaines ou des centaines de numéros.

En France et en Espagne, c'est avant tout une logique patrimoniale qui domine, parfois au sens large du terme avec l'inclusion des études critiques. Celle-ci est cependant remise en question depuis peu par l'arrivée des

1993, autant dire la préhistoire d'Internet. Malheureusement, le site semble en sommeil depuis 2002.

⁴<http://un2sg4.unige.ch/athena/html/francaut.html> Il s'agit d'une bibliothèque virtuelle proposée par l'Université de Genève. Le site est éditeur d'une partie des textes et son catalogue renvoie vers les liens adéquats pour les textes déjà édités par d'autres bibliothèques (BNF, Lisieux, ABU).

⁵<http://www.bmlisieux.com/> Le fonds est composé en grande partie de textes du 19^e siècle d'auteurs français, connus et moins connus.

⁶<http://www.cervantesvirtual.com/proyectoES/BIMICESA03.shtml>

bibliothèques virtuelles à but lucratif plus ou moins affiché. L'initiative Google Print a été accueillie comme une tentative d'hégémonie anglo-saxonne, un vieux démon de la pensée culturelle française. Cependant, comme le fait remarquer Jean-Michel Salaün, Professeur à l'ENSSIB (École Nationale Supérieure des Sciences de l'Information et des Bibliothèques) et responsable du réseau CNRS « Documents et contenus : création, indexation, navigation », cette initiative « n'est pas un projet hégémonique au sens culturel ou politique, mais fait partie d'une stratégie commerciale de l'entreprise américaine pour contrôler économiquement l'Internet. Le problème est avant tout économique et les seuls concurrents de Google sont actuellement Yahoo et Microsoft, porteurs eux aussi d'un projet de bibliothèque⁷ ». Le débat se retrouve donc placé sur le terrain économique, où il y a peu de chance qu'une institution publique vienne concurrencer des entreprises marchandes. Si on laisse de côté les problèmes patrimoniaux, du point de vue scientifique, le danger vient de logiques différentes car « si Google est transparent sur ses contenus, il reste opaque sur la technologie de son moteur de recherche. On ne sait pas vraiment comment il hiérarchise et indexe ses informations, et s'il favorise telle ou telle ressource sous contrat. C'est contre cette hégémonie commerciale et cette absence de transparence qu'il faut lutter⁸ ».

Ce rapide tour d'horizon permet déjà de deviner le premier problème qui se pose à l'hispaniste dans sa recherche sur Internet, celui de la dispersion des documents et de leur hiérarchisation.

1.2. La dispersion des documents et leur hiérarchisation

La dispersion des documents est devenue un fait majeur, étroitement lié à la multiplication des initiatives de numérisation et de mise en ligne. Schématiquement, l'augmentation du volume des documents mis en ligne est allée de pair avec leur dispersion. Cette situation a provoqué des réactions visant à juguler le phénomène. Il n'a pas été question de limiter les initiatives. Cette mesure était de toute façon pratiquement irréalisable et aurait eu un effet négatif sur le volume de documents disponibles.

⁷*Le journal du CNRS*, n°188, septembre 2005 <http://www2.cnrs.fr/presse/journal/2404.htm>

⁸Art. cit.

La solution avancée par les institutions publiques est passée par la mise en commun des catalogues et des chemins d'accès aux documents. Cette procédure permet de regrouper les références des documents dispersés et répond d'une certaine façon aux entreprises à but lucratif par une « coalition des institutions publiques ».

La *Biblioteca Virtual Miguel de Cervantes* (BVC) a ainsi signé des accords avec plusieurs institutions (universités, fondations, BNE) qui l'autorisent à recenser dans son catalogue les œuvres numériques qui appartiennent à leur propre fonds. Le catalogue de la BVC remplit ainsi la fonction de métacatalogue en facilitant l'accès à des documents dispersés en Espagne (dans le sens où ils ne sont pas directement hébergés sur les serveurs de la BVC). Mais à son tour, la bibliothèque collabore avec d'autres partenaires plus importants afin d'intégrer son catalogue à des catalogues de taille supérieure. Ainsi, après l'ajustement technique nécessaire, ses fiches sont maintenant accessibles sur le site de la bibliothèque du Congrès de Washington.

Parmi les projets en cours allant dans le même sens, on retiendra celui de Bibliothèque Méditerranéenne, piloté depuis Aix-en-Provence⁹. Il s'agit de « constituer un champ documentaire méditerranéen dans le cadre d'une démarche collaborative entre les membres du réseau [...] ». Un des intérêts de la démarche est qu'elle assemble les réalisations des partenaires, qui restent totalement autonomes dans la gestion de leur fonds et leurs projets de numérisation, mais chaque initiative est rendue compatible avec les autres par l'adoption d'un système de catalogage commun. Celui-ci réduira à terme la dispersion des documents, qu'il permettra d'atteindre par une requête unique, sans nécessité de consultations multiples.

En ce qui concerne la hiérarchisation, les initiatives institutionnelles font bien la distinction entre l'information scientifique, distincte de l'information d'opinion. La confusion des niveaux (commercial, scientifique, d'opinion...) de l'information disponible sur les réseaux est plutôt le fait d'initiatives privées, qu'elles soient commerciales ou individuelles. Cependant, le problème de la hiérarchie de l'information se pose de façon très visible lorsque le chercheur

⁹<http://bibmed.mmsb.univ-aix.fr/Presentation.html>

accède directement à l'Internet par un simple moteur généraliste (du type Google ou Altavista). Les résultats sont souvent brouillés par une information en provenance de niveaux non pertinents, bien qu'une utilisation pointue des paramètres de recherche avancée permette déjà un certain filtrage. Finalement, ces travers dénotent une utilisation bien plus intuitive que maîtrisée des outils et reflètent les carences de formation dans le domaine de la recherche de l'information¹⁰.

2. Formation et méthode

2.1. L'esprit de la formation : généralités et spécialisation disciplinaire

Dans l'état actuel des choses, Internet et les documents numériques auxquels il permet d'accéder ne remplacent pas le support papier, dont la maniabilité, la facilité d'usage et le prix sont des atouts incontournables (le coût d'une numérisation reste encore élevé)¹¹. Cette remarque vaut autant pour les bibliothèques numériques que pour les fonds d'archives. La formation à l'usage d'Internet dans le cadre d'une recherche documentaire ne doit donc pas ignorer les principes de la recherche papier, dont elle sera un complément indispensable. Par ailleurs, une formation généraliste est tout aussi nécessaire. Elle permettra d'aborder des éléments tels que les principes de fonctionnement et la logique d'un moteur (opérateurs booléens), l'évaluation de l'information, l'élaboration d'une requête, etc... Cependant, dans un second temps, on n'oubliera pas que, par des histoires et des traditions différentes, chaque discipline possède (ou ne possède encore pas) des outils et des ressources propres, comme c'est le cas pour le domaine hispanique. Après l'assimilation du socle commun interdisciplinaire, ce sont les spécificités disciplinaires qui devront être explorées et maîtrisées. Par exemple, les archives en ligne espagnoles, constituées par un énorme métacatalogue de 12 millions de notices renvoyant à des documents ou liasses de

¹⁰Voir à ce propos l'article d'Alexandre Serres, Moteurs de recherche et maîtrise de l'information : faut-il former à Google et comment ?, Communication présentée à Workshop Bucarest, *Le monde selon Google*, 2-4 juillet 2005. Texte accessible en ligne sur @archives SIC, http://archivesic.ccsd.cnrs.fr/sic_00001730.html

¹¹Voir à ce propos le rapport Van Dooren, publié par le Ministère de l'Éducation Nationale, de la Recherche et de la Technologie en juin 1999. <http://www.education.gouv.fr/rapport/vandooren/index3.htm#debut>

documents, n'ont aucun équivalent français à l'heure actuelle¹². Le fonctionnement de ce site, relativement complexe, ainsi que les documents et les fonds auxquels il renvoie devront être expliqués par un hispaniste ayant des connaissances « de terrain ». La formation doit donc aussi s'appuyer sur une connaissance disciplinaire, qui rendra moins théoriques les principes généraux et plus accessibles les ressources spécialisées.

2.2. La conceptualisation des outils et de la démarche de recherche

Pour partir d'une métaphore, Internet n'est qu'un support où le chercheur trouvera des outils qui lui permettront de construire une démarche de recherche cohérente. La cohérence reposera sur une organisation qui n'est pas sans rappeler la recherche documentaire classique (papier). Cette cohérence repose d'abord sur une identification précise des outils et de leur ordre d'utilisation. Par outils, nous entendons ici non pas un site spécifique mais une fonction. La nuance est importante, car il arrive souvent qu'un site remplisse plusieurs fonctions, proposant la gamme complète des outils nécessaires à la recherche. En essayant de conceptualiser les choses, on pourra distinguer 3 grands types d'outils, dont nous allons développer la définition et l'utilisation :

—Les catalogues et bases de données en ligne : ils permettent de repérer et localiser les documents.

—Les sites de ressources documentaires électroniques constitués par les bibliothèques, hémérothèques, et revues électroniques : ils permettent d'accéder à des documents en ligne.

—Les outils de veille, représentés par les listes d'alerte et les listes de distribution : ils permettent de se tenir informé de l'activité et la publication scientifique dans un domaine particulier.

Cette présentation ne suit pas un ordre aléatoire, elle correspond schématiquement à une démarche logique qui était celle de la recherche papier.

¹²AER. Archivos Españoles en la Red (<http://www.aer.es>). Ce métacatalogue est la somme des catalogues de différentes archives : Archivo Histórico Nacional, Archivo General de Indias, Archivo General de la Guerra Civil Española, Archivo General de Simancas, Archivo General de la Administración, Archivo de la Corona de Aragón, Archivo de la Real Chancillería de Valladolid, Sección Nobleza del Archivo Histórico Nacional Archivo Histórico Provincial de Álava. Ce gigantesque répertoire contient aussi des archives virtuelles et permet d'accéder directement à bon nombre de documents numérisés (classés par archives).

Internet rend cependant ce cheminement en partie théorique : la consultation du catalogue d'une bibliothèque virtuelle tend à réunir dans le même geste le repérage du document, son accès et sa consultation. Toutefois, il faut continuer à considérer ces deux étapes de façon différente, car ils ne correspondent pas au même stade d'une recherche dont nous proposons d'établir un schéma de fonctionnement.

3. Trois moments de la recherche

3.1. Première étape : catalogues, bases de données

Il s'agit du premier stade de la recherche, celui de la définition du sujet et de l'appréhension de ses limites. La recherche bibliographique permettra de localiser des documents qui serviront soit à l'établissement d'une bibliographie, soit à celui d'un corpus. Pour dresser cet inventaire, les catalogues et bases de données en ligne sont consultés prioritairement.

Pour une bibliographie, on consultera d'abord les trois grands sites que sont TESEO, l'ISBN, DIALNET. Par cette voie, on arrive à repérer assez largement les thèses, livres et articles écrits sur un sujet précis.

En Espagne, les outils sont pour le moment plus dispersés qu'en France. La base de données des thèses TESEO¹³ demande une consultation particulière, car elle n'est liée à aucun catalogue général, contrairement à la France où elle est englobée dans le système universitaire de documentation (SUDOC¹⁴). En revanche, contrairement à la France, l'accès à l'ISBN espagnol est gratuit. Cette base recense les ouvrages publiés depuis 1972 et constitue de fait une précieuse source d'information¹⁵.

Autre différence de taille par rapport au système français, l'existence d'un site de dépouillement de sommaires de revues : DIALNET¹⁶. Actuellement, la base recense 3 138 revues et 8 101 ouvrages collectifs, offrant au total une base de 812 011 références d'articles, qui devrait atteindre le million d'ici la fin de

¹³Cette base est assez semblable à la base française dans sa structuration. Elle rassemble les thèses soutenues depuis 1976. <http://www.mcu.es/TESEO/teseo.html>

¹⁴<http://www.sudoc.abes.fr/>

¹⁵<http://www.mcu.es/bases/spa/isbn/ISBN.html>

¹⁶<http://dialnet.unirioja.es/> Le site a été lancé en 2001 à l'initiative de l'université de la Rioja. En 2003, il est devenu un projet réunissant plusieurs universités du Nord et du Centre, ce qui a permis un développement plus rapide.

l'année. Ce site peut être utilisé d'abord comme une base de données. Le moteur de recherche proposé est pour le moment assez sommaire et ne permet pas de formuler des requêtes complexes. Cependant, ce n'est pas un simple outil bibliographique, DIALNET remplit aussi le rôle d'archive ouverte et de service de veille bibliographique, deux outils dont nous développerons l'utilisation plus loin.

Quelques ressources françaises sont également disponibles dans le domaine de l'hispanisme. Une consultation du SUDOC permettra de retrouver les thèses et ouvrages français sur un sujet donné. L'accès au site BIBLIOSH¹⁷, mis en ligne récemment par la Société des Hispanistes Français, permet d'accéder aux notices d'articles qui ne sont pas recensés par des bases de données hispaniques.

De façon complémentaire, on peut utiliser aussi les catalogues des bibliothèques virtuelles (voir infra) qui recensent des documents qui auraient pu échapper aux trois grands sites que nous venons d'évoquer.

Le travail de repérage bibliographique sera complété, le cas échéant, par la recherche de documents destinés à constituer un corpus. Les ressources sont alors très nombreuses et il est assez difficile de proposer des sites de référence car tout dépendra du type de corpus désiré. À titre d'exemple, nous retiendrons les grandes structures nationales. La BNE offre un catalogue en ligne, que l'on peut compléter par le métacatalogue du CCPBE (Catálogo Colectivo del Patrimonio Bibliográfico Español)¹⁸. Les fonds d'archives peuvent être scrutés grâce au site AER, évoqué antérieurement, qui englobe les catalogues des grands fonds d'archives espagnols. En linguistique, où la pratique s'est précocement informatisée, de nombreux corpus peuvent être constitués à partir de la presse en ligne, qui offre parfois plus de dix ans de journaux gratuitement. Une autre solution peut consister à recourir à des corpus en ligne, proposés avec des outils de recherche, comme le CORPUS DEL ESPAÑOL résultat d'un projet nord-américain soutenu par le NEH (National Endowment for the Humanities)¹⁹ ou

¹⁷<http://www.biblioshf.org>

¹⁸www.bne.es Catalogue de la Biblioteca Nacional de España.

¹⁹<http://www.corpusdelespanol.org> Le site, composé de 100 millions de mots, est proposé avec un moteur de recherche adapté aux préoccupations des linguistes. Il propose des échantillons allant du Moyen-Âge au 20^e siècle.

bien le CREA (Corpus de Referencia del Español Actual) ou le CORDE (CORpus Diacrónico del Español) proposés par la Real Academia²⁰.

2.2. Deuxième étape : consultation et lecture des documents en ligne

Les sites de ressources documentaires électroniques sont constitués par les bibliothèques, hémérothèques et revues virtuelles qui permettent d'accéder à des documents en ligne. Dans ce domaine, établir une liste de référence de l'hispanisme est un peu plus compliqué. En effet, les catalogues et les bases de données demandent un effort financier important et une certaine expérience historique dans le domaine, ce qui explique que ces sites soient peu nombreux, presque toujours institutionnels et assez fédérateurs. Par contre, dans le domaine de la mise en ligne des documents, les initiatives sont bien plus éparpillées. Malgré tout, on peut essayer d'envisager un itinéraire obligatoire, à compléter en fonction des spécialités internes à l'hispanisme. Prioritairement, on consultera la BVC qui possède le fonds le plus important d'ouvrages et d'articles en ligne. Le site DIALNET permettra éventuellement d'accéder au texte des articles. En marge de ces sites généralistes qui sont les plus fréquentés, on pourra, en fonction des besoins, s'intéresser à d'autres comme LOGOSFREEBOOKS, INTRATEXT²¹ ou aux bibliothèques des communautés autonomes qui proposent beaucoup de livres anciens²². Les sites spécialisés d'origine institutionnelle ou privée s'étant multipliés, on pourra accéder à des bibliothèques de poésie et de poésie, des collections de grammaires, de la presse ancienne, etc...²³. En dernier recours, le portail UNIVERSIA propose un répertoire de bibliothèques virtuelles et de grands projets de numérisation²⁴. En ce qui concerne les publications électroniques, leur éparpillement est la principale difficulté, encore que la

²⁰<http://corpus.rae.es/creanet.html>.

²¹<http://www.logosfreebooks.org> et <http://www.intratext.com/>

²²Sans être exhaustif, les sites suivants ont tous un intérêt :
<http://www.gipuzkoa.net/kultura/> BIBLIOTECA DIGITAL GUIPUZCOANA
<http://bdigital.ulpgc.es/mdc/> BIBLIOTECA DIGITAL GRAN CANARIA
<http://bv2.gva.es/default.php> BIBLIOTECA DIGITAL VALENCIANA
<http://www.euskadi.net/LiburutegiDigitala/> — BIBLIOTECA DIGITAL VASCA.
<http://www.juntadeandalucia.es/cultura/bibliotecavirtualandalucia/inicio/inicio.cmd> BIBLIOTECA VIRTUAL DE ANDALUCÍA.

²³Dans l'ordre de citation, pour la poésie <http://www.poesia-inter.net/index.htm> pour les grammaires et la stylistique <http://www.angelfire.com/de/hispania/> pour la presse ancienne <http://www.digibis.com/prensahis>

²⁴<http://www1.universia.net/CatalogaXXI/default.asp?IDC=10046&IDP=ES&IDI=1>

majorité des grandes revues électroniques espagnoles figure dans le catalogue de DIALNET. En complément, on pourra toujours consulter le portail UNIVERSIA, qui recense les adresses de revues par périmètre scientifique²⁵.

Au bout du compte, le problème principal de ces ressources est leur recensement incomplet. Des solutions sont en cours d'application comme les partenariats avec de grands sites, mais elles ne couvriront jamais l'ensemble de l'offre. Sur ce point, les chercheurs sont pour le moment condamnés à compléter leur liste de grandes bibliothèques par des listes personnelles qui pourront être mises à la disposition de la communauté universitaire. Le monde hispanique offre donc, par son dynamisme, un accès toujours plus important aux documents numériques, facilitant ainsi le travail du chercheur. Mais une autre révolution, plus lente, est en marche : celle de la lecture.

Avant de l'aborder un rappel sur les formats des textes est nécessaire. On peut, en simplifiant, en distinguer deux types : le document statique et le document dynamique, bien qu'une partie se situe un peu entre les deux.

Par document statique, nous entendons tout document numérisé dans un format qui le rapproche de l'image. Le fichier qui contient les données est généralement au format JPEG (extension .jpg) ou au format du texte universel (extension .pdf). Ces formats ont deux avantages. En amont, ils sont le résultat d'une numérisation « simple », l'équivalent d'une photocopie. Un livre ou un article peuvent ainsi être rapidement traités et mis en ligne grâce au gain de temps que suppose un format d'arrivée directement issu de la numérisation, sans aucun traitement postérieur du document. En aval, l'obtention d'une édition identique à l'édition d'origine présente un intérêt évident pour des ouvrages anciens et plus encore pour les manuscrits. On dispose donc d'un texte dont le mode de lecture sera identique à celui du papier, malgré la différence de support.

Un document dynamique se définira comme un support préalablement numérisé, mais qui aura subi un traitement avant sa mise en ligne, soit par une

²⁵<http://www1.universia.net/CatalogaXXI/default.asp?IDC=10046&IDP=ES&IDI=1> Ce portail est d'inégale valeur car il recense à la fois les revues électroniques et les revues papiers ayant une page sur Internet.

saisie manuelle, soit par le recours à l'OCR²⁶. Le résultat présente un avantage indéniable pour la lecture : souvent proposé au format XML, ce genre de texte peut être feuilleté grâce à des outils informatiques. Certaines bibliothèques comme la BVC proposent un moteur de recherche qui balaie l'ensemble du fonds qui répond à ce standard. L'outil permet de retrouver des citations, des noms propres, des phrases que la lecture traditionnelle aurait même peut être laissé échapper.

La seconde révolution dans la lecture provient de l'hypertexte²⁷. Reprenant une remarque du philosophe Pierre Levy, l'historien Marin Dacos fait remarquer que « cette cascade de liens compose une pensée collective originale, vivante, puissante et internationale dont il faudra bien un jour prendre la mesure. Notons au passage que cette innovation — issue des milieux universitaires — a conquis le public immédiatement car elle associe l'intérêt intellectuel de la référence à des qualités qui n'étaient pas présentes dans les notes de bas de page : la discrétion dans la page et la rapidité d'accès à la référence soulignée »²⁸.

Les nouveautés technologiques sont donc en train de changer en partie la pratique de la recherche. Le dernier secteur à évoquer est celui de la veille bibliographique.

3.3. Le travail de veille

Différentes sources d'information permettent, pour le chercheur confirmé, d'exercer, à son échelle et dans son domaine, une activité de veille. Sur Internet, l'activité de veille est facilitée car une partie des opérations peut être automatisée et personnalisée. L'outil du bulletin d'alerte sera sans nul doute le plus efficace sachant que les hispanistes possèdent l'immense avantage de disposer de ce service à travers le site DIALNET. En effet, après s'être enregistré gratuitement,

²⁶Acronyme anglais utilisé en français et qui désigne la reconnaissance optique des caractères sous forme logicielle.

²⁷Pour approfondir voir l'encyclopédie Wikipédia, à l'article suivant : <http://fr.wikipedia.org/wiki/Hypertexte> « Un système hypertexte est un système contenant des documents liés entre eux par des hyperliens permettant de passer automatiquement (en pratique grâce à l'informatique) du document consulté à un autre document lié. Un document hypertexte est donc un document qui contient des hyperliens ».

²⁸Voir à ce propos l'article de Marin Dacos, «Les lendemains électroniques de l'édition historique. Pour un nouveau modèle économique de publication périodique», *Revue d'histoire du XIXe siècle*, 2000 20/21, Varia, , mis en ligne le 20 juin 2005. URL : <http://rh19.revues.org/document218.html>. Consulté le 27 mars 2006.

le chercheur peut choisir de recevoir directement par courriel les nouveaux sommaires de toutes les revues dont il aura choisi au préalable la liste. Par ailleurs, ces revues étant indexées par mots-clés, lorsqu'un nouveau titre intègre la base de dépouillement, tous les abonnés à une liste contenant des revues du même domaine sont alertés et peuvent la rajouter à leur liste. Cette liste choisie par le chercheur, peut être modifiée à tous moments. L'activité de veille se situe ainsi au niveau le plus réactif qui est celui des publications périodiques.

Le deuxième outil, sans doute moins performant mais tout aussi intéressant, est représenté par les listes de diffusion. Le premier cas de figure est la liste de diffusion d'information fermée : une bibliothèque (par exemple la BVC) ou une librairie envoie les nouveautés de leurs catalogues aux abonnés qui reçoivent passivement le courrier. En revanche, la liste de diffusion ouverte est une structure créée par des membres qui communiquent entre eux. Ce sont les participants de la liste qui font l'information autour du thème qui les fédère. Ce service a été développé par le ministère de l'éducation en Espagne à travers le réseau REDIRIS qui favorise la diffusion scientifique²⁹. Généralement, l'information qui circule par courriel reprend l'annonce de colloques, de publications, mais aussi des questions sur une recherche en cours qui permettent souvent d'obtenir de l'aide de la part des co-participants. Beaucoup de domaines sont couverts. À titre d'exemple, on peut citer SIGLO XVIII qui rassemble des spécialistes du 18^e siècle, STYLUS qui propose de l'information sur la théorie littéraire et les ressources sur la littérature hispanique sur Internet, LAPEPA qui propose de regrouper les chercheurs travaillant sur la période contemporaine de l'histoire espagnole. Ces listes répondent au besoin de lutter contre l'isolement des chercheurs tout en favorisant les contacts hors des frontières.

4. En guise de conclusion, les limitations et les perspectives

Si on s'essaie à un bilan, forcément provisoire au cœur d'un phénomène en évolution, deux grandes lignes de force apparaissent. D'une part, les limites du « tout en ligne » et d'autre part une mutation dans la pratique de la recherche.

²⁹Voir le sommaire des listes sur <http://www.rediris.es/list/> où elles sont classées par domaine disciplinaire.

4.1. Les limites du « tout-en-ligne »

Ces limites qui se font jour sont de deux ordres. D'abord, tout chercheur, dans son expérience du terrain, a été confronté aux failles des catalogues qui ne peuvent tout référencer. Certaines archives demandent un travail... d'archiviste avant la réalisation du travail de chercheur. Le catalogue électronique ne peut ici être d'aucun secours.

Le second problème, beaucoup plus crucial sans doute, est celui de l'accès aux documents. La masse actuellement disponible en ligne ne représente qu'une part infime de la quantité documentaire exploitable. Comme nous le disions au début de l'article, le support numérique a un coût élevé. Certaines opérations, comme l'intervention humaine lors de la procédure de numérisation et éventuellement d'OCR, sont absolument indispensables et irremplaçables dans l'état actuel de la technologie. Le travail sur le support papier garde encore de beaux jours devant lui. La vraie révolution, plus que la lente évolution que nous venons d'analyser, sans doute faut-il la chercher du côté de la publication scientifique électronique.

4.2. Vers une mutation épistémologique

Le cas des revues électroniques mériterait à lui seul une étude, mais nous allons cependant essayer d'étudier les principaux éléments de cette mutation.

Fondamentalement, il n'y a pas de différence entre le support papier et le support électronique. Le développement de ce dernier répond à des facteurs historiques et techniques que nous n'avons pas le temps d'analyser ici. On remarquera que la revue électronique est assez répandue dans le monde hispanique alors qu'en France, la majorité des enquêtes évoquent un « retard provisoire » en la matière. La quasi-totalité des universités espagnoles possède plusieurs revues en ligne et offre l'accès, par un archivage rétroactif, à des publications papier antérieures. À moyen terme, on semble se diriger vers du tout électronique, et cela pour deux raisons. D'abord la raison économique: une revue électronique n'a pas de coût d'impression, de diffusion et de stockage. Ensuite une raison scientifique : la diffusion et le rayonnement d'une revue électronique (principalement en raison du support) sont sans commune mesure avec le papier. La garantie de la validité scientifique reste la même, assurée par le

fonctionnement des comités de lecture sur lequel le support n'a aucune incidence. De façon marginale ensuite, le support électronique permet d'éviter la pensée « calibrée » ou l'exigence « de l'article de 15 pages », souvent vécue comme une norme répressive. Une exigence finalement beaucoup plus matérielle que scientifique, car conditionnée par la taille du livre. Par ailleurs, l'édition électronique permet de nouvelles formes d'édition : la publication de sources et de documents complexes soutenant l'argumentation des auteurs devient possible (documents audio ou textes d'archives). L'augmentation de l'information en ligne permet d'accorder de plus en plus de place à l'hypertexte qui rapproche le lecteur des sources directes de l'article qu'il est en train de consulter.

L'ensemble des pratiques génère une nouvelle façon de lire, de penser et d'écrire dont les implications sont encore mal cernées et sans doute en pleine évolution.

Marc MARTI

CIRCLES, Université de Nice