

## **La revitalisation numérique du patrimoine littéraire territorialisé**

Jean CASENAVE  
Laboratoire IKER – UMR 5478, CNRS – Bordeaux 3 – UPPA

Christophe MARQUESUZAÀ  
Laboratoire LIUPPA - Équipe d'accueil n° 3000 - IUT de Bayonne –  
Pays Basque

Pantxika DAGORRET  
Laboratoire LIUPPA - Équipe d'accueil n° 3000 - IUT de Bayonne –  
Pays Basque

Gaio MAURO  
Laboratoire LIUPPA – Équipe d'accueil n° 3000 - UFR Sciences et  
techniques - Pau

Dans tous les pays de l'Europe de l'ouest, les institutions nationales<sup>1</sup> prennent des initiatives pour mettre leur patrimoine culturel d'intérêt national ou universel en valeur par le biais des technologies numériques et du web. En France, le Musée du Louvre, la Bibliothèque nationale de France et son serveur électronique Gallica, l'Institut National de l'Audiovisuel et l'ensemble des grands établissements offrent des services variés qui vont de la visite virtuelle à la mise en ligne de contenus téléchargeables ou consultables, en passant par une gamme étendue d'activités adaptées aux divers publics potentiels.

En revanche, la valorisation des fonds de notoriété moindre est beaucoup plus ardue, qu'il s'agisse de corpus relevant de registres « exigeants » ou plus circonscrits en termes de rayonnement géographique. Pour s'en tenir à l'exemple français, on peut toutefois citer quelques tentatives intéressantes, à l'image de celles que proposent la bibliothèque électronique municipale de Lisieux (<http://bmlisieux.com>) et le service des Archives de la Région Savoie sur son serveur électronique *Saubadia* (<http://www.saubadia.org>). Cependant, le domaine demande à être développé car le patrimoine dont disposent ces espaces territoriaux est aussi riche que varié, et son potentiel de valorisation tout à fait incontestable.

---

<sup>1</sup> Bibliothèque nationale de France : <<http://gallica.bnf.fr>> - Archives littéraires suisses : <<http://www.nb.admin.ch>> - Bibliothèque nationale du Luxembourg, Centre national de littérature : <<http://literaturarchiv.lu/fr/>> - Bibliothèque royale de Belgique : <<http://www.aml.cfwb.be>> [février 2007].

C'est pourquoi, dans le cadre de ce colloque consacré au cycle de vie du document, nous proposons une réflexion sur la mise en valeur des fonds territorialisés – locaux, régionaux – et tout particulièrement des ensembles documentaires constitués autour des corpus littéraires anciens. Nous étudierons les conditions de « revitalisation » culturelle et technique d'un corpus littéraire « oublié ». Nous nous efforcerons de montrer que, dans ce nouveau « marché des biens culturels » ouvert à la concurrence, il y a bien une place pour l'expression des singularités régionales dès lors qu'il existe un projet de socialisation conçu au plus près des ressources numérisées, dans un esprit de coopération entre les instances en charge du développement des territoires et les équipes de recherche spécialisées dans les contenus et les technologies de la communication.

Nous traiterons d'abord de l'intérêt patrimonial de ces fonds documentaires, pour aborder ensuite la liaison entre la conversion numérique des fonds, la problématique de leur diffusion et celle des dispositifs techniques et conceptuels qui président à leur communication. Par leur ancrage dans les lieux qu'ils évoquent, ils livrent d'indispensables éléments de connaissance objective mais aussi de précieuses représentations d'ordre subjectif. Indubitablement, ils contribuent à enrichir ce nouvel imaginaire des territoires que les nouveaux récits pédagogiques, médiatiques ou touristiques s'efforcent de reconstituer, bien trop souvent dans l'ignorance des sources bibliographiques disponibles. À l'heure de l'élargissement de l'Europe, ce repositionnement de « l'archive territorialisée » au sein du discours culturel, représente un enjeu important. Par leur conversion au format électronique, ces ressources deviennent l'objet d'un partage élargi de l'information et des savoirs qui va bien au-delà des communautés scientifiques et éducatives. Leur socialisation touche à l'espace public dès lors qu'elle entre dans le champ du développement économique régional par le biais des activités touristiques ou par son ouverture au grand public.

Ce sont donc là autant de nouveaux usages à l'adresse de nouveaux publics, et les technologies de l'informatique et des multimédias permettent d'assurer ce polymorphisme au document traité. Pour éclairer ce point technique dans la deuxième partie de notre exposé, nous aborderons la question de la préparation conceptuelle et technique de ces documents.

Qu'il s'agisse de littérature populaire, savante ou encore de représentations iconographiques, l'omniprésence des lieux caractérise le fonds local ou territorialisé et les documents fourmillent de notations et de repères spatiaux (onomastique, séquences descriptives, etc.). Compte tenu de la spécificité de l'archive territorialisée, nous avons choisi d'expérimenter une didactisation à partir de la sémiotique du territoire. Pour être véritablement opératoire, la préparation informatique du document (pose de balises ; appareillage conceptuel, technique et ergonomique) doit s'opérer dès la numérisation. En fonction de paramètres tels que le genre du « document-source » ou celui des publics ciblés, la préparation de l'application numérique devrait pouvoir se faire selon des

protocoles reproductibles et extensibles à un large éventail de ressources relevant des corpus territorialisés.

L'équipe de recherche interdisciplinaire<sup>2</sup> que nous avons constituée pour mener à bien cette étude de la didactisation des corpus territorialisés (textes littéraires, ressources iconographiques et phonographiques) se compose de chercheurs spécialisés dans les domaines de la langue et de la littérature, de l'informatique et du multimédia, et elle pourrait prochainement recevoir l'appui de spécialistes de l'aménagement du territoire et de la gestion de projet.

## Conditions de resocialisation d'un corpus restreint

### La question des « corpus restreints »

#### Une problématique spécifique, des contraintes économiques

Compte tenu des coûts élevés que représente le traitement informatique des fonds « papier », les corpus restreints ou d'intérêt moindre au plan du rayonnement géographique, linguistique, littéraire ou culturel, ne peuvent pas bénéficier des mêmes possibilités de numérisation que les œuvres perçues comme majeures : ils ne possèdent pas la même « légitimité culturelle » auprès du public ni la même reconnaissance institutionnelle. La transposition électronique de l'ouvrage d'un auteur considéré comme secondaire ou celui d'un romancier en voie de consécration sera difficile à justifier auprès d'un grand établissement de diffusion.

Il en est de même pour des genres moins prisés aujourd'hui comme la poésie (à l'exception des valeurs sûres) ou encore la production poétique, théâtrale ou journalistique de certains écrivains consacrés des XVIII<sup>e</sup> et XIX<sup>e</sup> siècles. En effet, ce sont toujours les mêmes titres (romans, essais, recueils de poèmes, d'articles ou pièces de théâtre en fonction de l'image que notre époque se construit de cet auteur) qui profiteront des possibilités de publication et, par voie de conséquence, auront quelque chance d'accéder à la numérisation. À ces cas particuliers, il faut ajouter d'autres corpus restreints qui sont écrits dans des langues rares et peu répandues ou qui présentent une

---

<sup>2</sup> Cette équipe est constituée par des chercheurs appartenant aux laboratoires indiqués ci-dessous : IKER est une Unité mixte de recherche (5478) qui regroupe des membres du CNRS et des enseignants chercheurs de l'université de Bordeaux 3 et de l'université de Pau et des Pays de l'Adour. Centre de recherche sur la langue et les textes basques, ce laboratoire travaille sur le domaine basque à partir de directions de recherche variées. LIUPPA est le laboratoire d'informatique de l'université de Pau et Pays de l'Adour, Équipe d'Accueil n°3000. Les auteurs de cet article appartiennent à l'équipe IDEE (Interaction, Document Électronique et Éducation). Au cours de l'année 2003, ces deux équipes ont décidé de s'associer autour d'un thème de recherche commun : la valorisation du patrimoine local dans la culture numérique. L'étude proposée à l'occasion de ce colloque a été menée dans le cadre de ces activités communes.

spécificité géographique. À titre d'exemple, nous développerons ici la problématique qui est attachée aux fonds locaux et territorialisés.

#### **Un exemple de corpus restreint : le corpus « local » ou territorialisé**

Toute bibliothèque, en ville comme dans les bourgs ruraux, réserve une partie de sa surface d'exposition aux livres qui évoquent la ville, la région, le terroir, bref, un territoire plus ou moins bien identifié, sujet de prédilection d'un ensemble de publications que les spécialistes s'accordent à désigner par l'expression « fonds local ». Nous l'appellerons plutôt « territorialisé » pour éviter le caractère restrictif et parfois condescendant qui est attaché au terme « local » par opposition à « national » ou « universel ». Le qualificatif « territorialisé » indique aussi que ce corpus s'inscrit explicitement dans cet espace déterminé, il s'en inspire et s'en imprègne, il en explore les dimensions physiques et oniriques, déploie son imaginaire et le met en scène avec bien souvent un luxe de détails qui fait la spécificité de ce type de document.

Aux ouvrages présentés au public, il faut généralement ajouter le nombre considérable des livres qui figurent dans les réserves et qui, n'apparaissant que sous la forme d'une mention dans le catalogue de l'établissement, ont bien peu de chances d'attirer l'attention du lecteur ordinaire. Au côté d'ouvrages très divers (histoire locale, géographie locale, économie, traditions, « beaux livres » essentiellement photographiques, etc.), la littérature (contes et littérature populaire écrite et orale, récits de voyages, romans ruraux et régionalistes, iconographie du territoire) occupe une place de choix. Prisés par un public toujours renouvelé (bien que relativement restreint), de curieux, d'érudits, d'universitaires ou d'étudiants, les ouvrages littéraires qui garnissent ces rayons « régionalistes » ou « territorialisés » ont *a priori* peu de chances d'accéder au nombre des rééditions. Comme nous le verrons plus loin, ces divers points lui confèrent un caractère exemplaire et constituent un bon potentiel de « revitalisation ».

#### **Un champ d'expérimentation interdisciplinaire**

Cette réflexion à laquelle nous avons souhaité donner le caractère le plus général possible s'appuie sur une expérience interdisciplinaire bien précise. En effet, la collaboration entre les deux laboratoires cités plus haut (IKER-LIUPPA) s'est concrétisée – à l'occasion d'un appel à projets lancé par les institutions locales (département, région) – auquel nous avons répondu en proposant le thème de la valorisation numérique du patrimoine local. Nous avons engagé des recherches dans diverses directions, et celle de la resocialisation numérique des fonds bibliographiques, iconographiques et phonographiques du domaine basque relatif à la première moitié du XX<sup>e</sup> siècle nous est apparue à la fois riche au plan documentaire, et très prometteuse en termes d'interdisciplinarité. Cela nous a conduits dans un premier temps à définir les bases d'un traitement

interdisciplinaire des documents puis, en fonction de ces critères à la fois théoriques et techniques (corpus « oublié », redoublé par une documentation iconographique et phonographique, etc.), à proposer la numérisation d'un « corpus aux contours bien définis » afin d'étudier les possibilités de valorisation informatisée qu'il recelait.

En conséquence, dans le cadre d'une collaboration étroite entre spécialistes des documents (textes et des contenus iconographiques et phonographiques) et informaticiens, nous avons sélectionné un ensemble de textes issus de la littérature populaire du XIX<sup>e</sup> siècle (contes, proverbes, chansons) et un petit nombre de romans en langues basque et française<sup>3</sup> qui décrivent tous abondamment le Pays basque. Ces derniers ont en commun d'avoir tous été publiés entre 1893 et 1946 et appartiennent donc au courant littéraire régionaliste qui a connu son heure de gloire dans la première moitié du XX<sup>e</sup> siècle. Ils présentent tous la caractéristique d'avoir connu une certaine notoriété à l'époque de leur parution mais n'ont pas été republiés depuis la fin de la seconde guerre mondiale.

La remise en circulation de ces fonds documentaires représente un défi technique et économique intéressant ainsi qu'un enjeu culturel important. Les propositions qui résultent de cette réflexion interdisciplinaire sont autant d'hypothèses de travail, sans doute partagées par la plupart des équipes de chercheurs qui étudient la valorisation numérique d'un « corpus restreint ». C'est ce qui nous a amenés à penser que les problématiques auxquelles nous nous sommes heurtés à propos de notre corpus territorialisé étaient, dans une large mesure, transposables à la plupart des cas de figure présentés par les corpus restreints.

### **Les caractéristiques du corpus territorialisé**

Chaque corpus restreint possède à la fois ses atouts et ses faiblesses. Le recensement de ces éléments et leur étude interdisciplinaire sont fondamentaux dans la perspective de la valorisation numérique du corpus. Comme on le verra ci-dessous, les points forts du corpus, l'ancrage territorial en l'occurrence, donneront prise sur les documents et constitueront l'axe principal du travail interdisciplinaire.

#### **Un fonds inscrit dans un ensemble historico-géographique**

Dans la tradition littéraire française, G. Sand pour les terres du centre de la France, J. Giono pour la Provence, E. Le Roy pour le Périgord, H. Pourrat pour l'Auvergne, P. Loti pour le Pays basque et la Bretagne viennent immédiatement à l'esprit lorsqu'il s'agit d'associer une œuvre à un territoire. Et il y a les innombrables ouvrages d'auteurs oubliés qui ne demeurent vivants que pour

---

<sup>3</sup> En matière de fiction, la plus grande part de la création de l'entre-deux-guerres est d'expression française.

quelques dizaines de lecteurs « spécialisés ». Plus ou moins longuement après leur parution, ces livres ont connu une notoriété nationale, régionale ou seulement locale. Puis, non réédités et tombés dans l'oubli, ils dorment dans les réserves des bibliothèques et ne trouvent un délai de grâce que chez les bouquinistes où les habitués sont souvent davantage collectionneurs que lecteurs.

Cependant, ces fonds constituent des sources d'informations précieuses à de nombreux titres. Inscrits dans un ensemble historico-géographique déterminé, ils font référence de façon dense et précise à des lieux qui sont nommés de manière explicite : toponymes bien sûr, mais aussi noms de points d'eau ou de rivières, noms de villages ou de villes, noms de maisons aussi, dans le domaine basque très souvent en étroite relation avec les noms personnages qui sont également des domonymes. Ces documents abondent en descriptions particulièrement précises du territoire et des paysages mais évoquent aussi les mœurs et des activités laborieuses comme festives.

Le corpus des cartes postales, celui des images en général (photographiques, cinématographiques et vidéographiques), d'une manière souvent redondante par rapport à l'écrit, est à son tour venu « consacrer » les mêmes paysages (le rivage, la montagne, la forêt, la ferme, etc.), les mêmes activités ludiques (la partie de pelote, la danse, la pastorale ou la foire, etc.) ou les mêmes personnages (le chanteur, le berger, le bouvier, le grand-père, etc.), jusqu'à en faire des « représentations stéréotypées ». À l'intérieur comme pour l'extérieur, c'est donc l'image d'un « terroir » aujourd'hui « vécu<sup>4</sup> » comme territoire qui s'est constituée autour des représentations proposées par ce fonds « papier » et par ses adjuvants. Bien que remplacé par les nouveaux discours du territoire dont il sera question plus loin, on peut déjà avancer l'idée que l'imaginaire local contemporain est encore largement influencé par cette source, et ceci sans le savoir et sans y faire directement référence, simplement en s'appuyant sur les motifs stéréotypés qui ont perduré et qui sont invariablement repris avec peu de recul critique.

À un moment où la notion de « territoire<sup>5</sup> » relance l'ancien concept de « terroir » d'une façon largement renouvelée et ajustée à nos préoccupations actuelles par la géographie humaine et l'anthropologie contemporaines, toutes les évocations localisées retrouvent un intérêt, qu'elles soient textuelles, iconographiques ou sonores. C'est sous cet éclairage qu'il est possible d'opérer une relecture et une réévaluation de ces fonds dans le champ culturel contemporain. De plus,

---

<sup>4</sup> Cette expression fait bien sûr référence au livre du géographe (Frémont, 1976), ouvrage pionnier en la matière mais aussi à son étude sur la Normandie telle qu'elle apparaît dans les récits de Maupassant.

<sup>5</sup> Dans le cadre de cet article, il n'est pas possible d'intégrer des développements nourris de l'apport effectué sur ces questions par les géographes, les historiens, les anthropologues, les ethnologues et les spécialistes des sciences de l'information et de la communication (cf. Bibliographie).

l'informatisation des corpus documentaires renouvellent sérieusement les opportunités de resocialisation et de valorisation à leur endroit.

### **Un fonds « oublié » et invisible**

Cependant, avant d'envisager leur « renaissance », il faut savoir ce qu'il en est, à l'heure actuelle, de l'existence des corpus locaux ou territorialisés des périodes citées. Pour l'essentiel, ils sont devenus des documents d'archives et ne connaissent plus désormais que la fréquentation d'un public très réduit et spécialisé qui en fait une utilisation savante (recherche universitaire, érudition locale, source de références et dictionnaire de citations – souvent les mêmes – pour les journalistes).

Oublié et remisé dans les magasins des bibliothèques, ce fonds souffre donc d'un manque de visibilité. Il faut ajouter à cela, un indéniable déficit de lisibilité pour le non spécialiste. La difficulté, comme pour la plupart des textes littéraires ou documentaires du passé, tient bien sûr à l'obstacle linguistique et stylistique, c'est-à-dire à un état de langue et de style littéraires désormais très daté.

Mais elle provient également des contenus eux-mêmes car le discours tenu sur « l'espace local » peut sembler lui aussi très daté puisqu'il participe d'un mouvement de « singularisation » à outrance des « terroirs » bien caractéristique des courants folkloristes et régionalistes de la première moitié du XX<sup>e</sup> siècle mais difficile – voire impossible – à comprendre pour les lecteurs contemporains.

Autant dire que la remise à disposition du grand public de ce corpus nécessiterait un important travail d'historicisation, notamment à travers la mise au point d'un appareil critique. Un premier pas vers la revitalisation de ce corpus consisterait à le répertorier de façon systématique dans le catalogue électronique des établissements, ce qui n'est pas le cas pour l'instant, puisque, pour la plupart, les œuvres ne figurent que dans le catalogue « papier ». Cela leur permettrait d'accéder au statut « d'archive littéraire » et les rendrait visibles et, par voie de conséquence, consultables ne serait-ce que sous sa forme « papier ». Cependant, il est indiscutable que ces fonds possèdent d'autres atouts de revitalisation.

### **Le potentiel de revitalisation de ce fonds**

Comme indiqué plus haut, ce corpus porte en lui une connaissance approfondie des lieux et du territoire qu'il évoque. Ce point constitue le cœur de son potentiel de revitalisation. Malgré son caractère daté, il continue à faire sens en termes de données géographiques, historiques, ethnographiques et littéraires par exemple. Les nouveaux récits pédagogiques, médiatiques ou



touristiques du territoire s'efforcent de reconstituer ces savoirs et, sans nul doute, ils gagneraient à se ressourcer dans les œuvres originales pour étoffer leurs descriptions et les appuyer de citations plus variées.

Paradoxalement, c'est même son caractère « daté » qui en fait le prix. En effet, il représente une mémoire de ces espaces et des visions dont ils ont fait l'objet dans un passé somme toute encore récent. Inscrits dans des représentations concrètes du territoire ou relevés comme les affleurements d'un imaginaire et d'une subjectivité plus intimistes, ces repères littéraires ou culturels situés au cœur des documents sont susceptibles d'alimenter ce « grand récit du territoire » qui se développe dans la littérature contemporaine ou dans la presse spécialisée (magazines consacrés à des territoires spécifiques ; exemples : *Alpes magazine*, *Pyrénées Magazine*, *Pays Basque Magazine*, *Bretagne Magazine*, etc.), que dans les publications les plus variées (anthologies, albums photos, nombreuses publications liées à la découverte de l'environnement, etc.), dans le matériel pédagogique préparé à l'intention des classes par les services éducatifs spécialisés et dans les manifestations culturelles et touristiques.

Bien entendu, la lisibilité du corpus doit être assurée par un appareil critique adéquat. Les documents cités participent donc pleinement d'une sémiotique du territoire qui, comme on le verra plus avant, peut devenir le dénominateur commun pour les utilisateurs comme le concept de traitement des données des concepteurs de la didactisation.

Les raisons *a priori* objectives de la remise en circulation de ce corpus paraissent donc bien établies ; encore faut-il voir maintenant à quelles conditions il est possible de réintroduire ces données dans le flux des savoirs et des informations qui circulent dans la société contemporaine. C'est aussi l'occasion d'observer comment elles peuvent s'intégrer dans la culture numérique en cours de constitution.

### **La resocialisation du corpus**

Si l'on envisage de resocialiser un corpus restreint, il faut y trouver des entrées qui conviennent aux utilisateurs actuels « réels » c'est-à-dire déjà actifs comme potentiels (extension du nombre et des types d'utilisateurs). Il s'agit donc de trouver des dénominateurs communs qui rapprochent les diverses communautés.



### **Numérisation et « réalisme économique »**

La numérisation représente pour ces fonds documentaires (bibliographiques, iconographiques, phonographiques) oubliés, la seule opportunité d'ensemble véritablement réaliste d'une resocialisation et d'une « revitalisation ». En effet, il n'est pas envisageable de faire republier un ensemble de textes désormais difficilement « lisibles ». Certes, quelques maisons d'éditions ont tourné la difficulté en publiant des morceaux choisis ou en constituant des ouvrages thématiques ou encore en proposant des textes peu connus d'auteurs célèbres (Victor Hugo). Toutefois, il est impensable de mettre en place une politique éditoriale « papier » qui permettrait une resocialisation d'un corpus étendu.

En revanche, la transposition numérique peut offrir de nouvelles perspectives. En effet, la conversion au format électronique modifie grandement les modes d'utilisation traditionnels puisque sa transformation en fichier informatisé scelle son polymorphisme et sa « banalisation » parmi bien d'autres ressources possibles (iconographiques et sonores : fonds photographiques, cartographiques, cinématographiques, etc.). Dès lors, ces documents numérisés deviennent utilisables pour eux-mêmes avec toutes les commodités que procure le format électronique : consultation à distance, téléchargement, annotations, etc. Ils peuvent aussi faire l'objet d'une valorisation en association avec les autres ressources évoquées plus haut, elles aussi informatisées.

Encore faut-il que les modalités de diffusion de ces ressources soient trouvées. Cela implique la nécessité de concevoir un projet de valorisation des corpus numérisés au-delà du seul souci de préservation d'un héritage du passé, sous la forme d'archives.

### **Le repositionnement du « corpus local » dans le champ culturel**

Il s'agit donc de rendre les contenus de ce corpus au public, à travers un usage élargi et actualisé dans le cadre de la culture numérique. En quelque sorte, il faut le repositionner parmi les ressources actuellement disponibles à propos du territoire évoqué et lui donner un statut nouveau. Ce processus de resocialisation a deux aspects complémentaires, un versant technique et un versant conceptuel et ils sont tous les deux étroitement imbriqués.

Le changement de statut de ce fonds est complexe. Il doit retrouver une pertinence qu'il a perdue au milieu des nouveaux récits du territoire évoqués. Comme indiqué plus haut, ceux-ci proviennent de milieux différents mais ces discours aussi bien journalistiques, politiques, économiques, éducatifs que touristiques, s'appuient tous sur la même entité territoriale et partagent avec le discours littéraire du passé la même volonté de mettre cet espace plus ou moins défini en valeur pour asseoir leur légitimité.

La resocialisation passe nécessairement par deux opérations qui sont grandement facilitées par la transposition des œuvres en version électronique. En effet, grâce à la numérisation, l'œuvre devient visible. Par le simple fait de la faire apparaître dans un catalogue électronique, la ressource reprend vie. Dès lors, elle peut circuler à titre de référence. On peut aussi envisager sa mise à disposition totale auprès du public puisque la numérisation a transformé l'œuvre en fichier informatique. Il est donc possible d'afficher le texte, de le télécharger, de le manipuler au sens propre du terme, de l'annoter, de le fractionner pour l'utiliser à des fins particulières. Mieux, on peut aussi penser à le transformer avant de le proposer à nouveau à la lecture. Bref, après l'avoir rendu à nouveau visible, le processus de numérisation peut lui rendre une « existence » socialisée en le remettant en circulation à travers les voies largement ouvertes de l'Internet. Ainsi, dans le cadre de la culture numérique, la « ressource territorialisée » produite localement qui n'avait qu'un rayonnement limité, se voit proposée à un public aléatoire et considérablement élargi, tout au moins potentiellement.

Enfin, cette visibilité renouvelée, cette remise en circulation décuplée par les moyens techniques contemporains, propulse ce corpus « oublié » au rang d'archive littéraire répertoriée et utilisable par le plus grand nombre. De plus, elle transforme insensiblement ce fonds d'archives en patrimoine littéraire territorial. Ce nouveau statut d'objet patrimonial classe ces œuvres du passé au rang de « monument » à la fois source de témoignages dignes d'intérêt et de « lieu » (virtuel) qui mérite le détour pour une visite.

On voit sans peine que la numérisation a rendu le corpus considéré à sa vocation de référence tout en le remettant au cœur des activités culturelles ; en un mot, elle l'a redynamisé, revitalisé. Toutefois, dans le même temps, elle a modifié le rapport à ces ressources, leur utilisation et, par voie de conséquence, leur interprétation. On peut penser qu'au prix de ces transformations, ces documents ont été rendus à leur public potentiel à savoir, les universitaires et leurs étudiants, les érudits, les professionnels de la plume, écrivains et journalistes, ceci, dans une forme actualisée qui facilite grandement l'utilisation des documents. Pourtant, il ne fait guère de doute que les divers changements évoqués ouvrent de nouvelles perspectives aussi bien en matière d'usages nouveaux que d'élargissement du public susceptible de fréquenter et d'utiliser ce corpus.

#### **Nouveaux publics, nouveaux usages**

À l'occasion de nos recherches interdisciplinaires sur la numérisation du patrimoine littéraire attaché au domaine basque, nous avons travaillé sur la question de la resocialisation du

corpus tant du point de vue des spécialistes des contenus que des concepteurs de l'équipement numérique des applications. Tout en sachant que la mise à disposition du corpus sous la forme d'application informatique en circuit contraint et surtout sur l'Internet, donne à l'éventail des publics et des utilisations possibles un caractère aléatoire et imprévisible, nous avons privilégié trois types de publics pour trois types d'usages.

Il s'agit de la communauté scientifique et « érudite<sup>6</sup> » pour un usage savant, puis de la communauté scolaire pour un usage pédagogique et enfin, du grand public susceptible d'avoir un usage informativo-ludique ou informativo-documentaire du corpus, par exemple dans le cadre d'une exploitation touristique des ressources numérisées.

Il est indéniable que l'usage le plus attendu correspond à celui de la communauté scientifique puisque, qualifié ici de « savant », il se place dans le droit fil de l'utilisation traditionnelle de ce corpus territorialisé, à savoir, une étude approfondie du texte, de ces sources et du contexte qui l'a vu naître. Cette valorisation du document débouche souvent sur la réalisation d'une édition savante dotée d'un appareil critique qui en enrichit la lecture à défaut de toujours la faciliter.

Dans la mesure où il existe au format électronique, ce corpus pourrait aussi donner lieu à une préparation pour une utilisation fragmentaire, éventuellement en association avec d'autres documents disponibles dans d'autres médias. Cette présentation peut intéresser de nouveaux publics qui ne feront pas du fonds une utilisation exhaustive à l'instar des utilisateurs traditionnels. Ainsi, c'est la perspective de toucher de nouveaux publics susceptibles d'expérimenter de nouveaux usages des documents qui découle de la transposition numérique.

Pour être en mesure d'aborder la question du partage des corpus avec de nouveaux publics, il faut auparavant évoquer les changements profonds que l'informatisation entraîne dans le rapport au texte et à l'œuvre littéraires. Dès lors qu'il est numérisé, le texte littéraire peut éventuellement être intégré dans une base de données extérieure à son environnement habituel et, dès lors, se pose la question de l'élargissement des publics susceptibles de s'intéresser à lui. En effet, il ne s'agit pas simplement de vouloir numériser des corpus parce que la possibilité technique existe. Il faut aussi savoir pourquoi et pour qui le faire. Il faut donc avoir un projet pour ce patrimoine, un projet de transmission, de partage. Il faut alors réfléchir aux conditions de cette ouverture, de ce partage.

---

<sup>6</sup> L'expression « communauté érudite » recouvre ici le nombre des lecteurs spécialisés qui sont habituellement qualifiés « d'érudits locaux ».

Pour se limiter au patrimoine littéraire territorialisé et sans chercher le moins du monde l'exhaustivité, nous nous contenterons d'illustrer cet élargissement des publics et des usages à travers deux exemples situés aux deux extrêmes du spectre des activités rendues possibles par l'informatisation des ressources : la communauté éducative et les professionnels du tourisme.

Le texte peut fort bien être reçu de façon autonome comme celui de toute œuvre littéraire mais son ancrage territorial peut également lui permettre de constituer un précieux apport d'informations dans d'autres modalités de lecture. Qu'ils servent de repères, de descripteurs explicites ou même de caution culturelle sous forme de citations ou d'allusions dans un article de revue, voire dans une brochure touristique, ces éléments prennent tout naturellement place au rang de références indispensables et parlantes.

Ils peuvent aussi de façon moins parcellaire être inclus dans la documentation pédagogique d'une classe-patrimoine, par exemple parmi des ressources de types et d'origines très divers : collections photographiques, cinématographiques, vidéographiques, phonographiques, etc. En effet, ces repères linguistiques et littéraires doivent s'intégrer parmi d'autres sources de connaissance du territoire, qu'il s'agisse des données liées à la nature (géologie, hydrologie, flore, faune, etc.), ou des études scientifiques au sujet des activités humaines traditionnelles ou contemporaines.

Depuis quelques années, de nouveaux lieux réels ou virtuels sont aménagés pour exposer ou utiliser les matériaux et les objets qui permettent d'accéder à la connaissance des territoires et ils sont souvent mis en œuvre par les institutions locales (département, région) : musées, centres culturels, services pédagogiques des musées, classes-patrimoines, etc. À ces lieux bien réels qui utilisent les technologies numériques dans leurs activités, il faut joindre par exemple, les expositions itinérantes ou permanentes qui sous la forme de bornes interactives, mettent la culture locale à la disposition du grand public. À cela, il faudrait encore ajouter les visites virtuelles des musées, les bornes consultables dans les offices de tourisme ou les sites web et autres portails de la culture qui intègrent le domaine territorialisé et qui touchent aussi bien le grand public local que celui du tourisme.

À travers l'évocation de cet élargissement des publics susceptibles de s'intéresser aux fonds locaux, il apparaît clairement que les mentalités vont devoir évoluer de façon notable. Les littéraires et autres spécialistes des contenus ne vont plus pouvoir se tenir fermement sur leurs principes quant à l'intégrité du texte littéraire ou de l'œuvre d'art. Les méthodes de travail des informaticiens et des experts en communication vont nécessairement bousculer les habitudes de leurs homologues linguistes ou géographes, et inversement. Pour être en mesure de proposer la valorisation numérique

d'un corpus, il faut lui appliquer un traitement que de nombreux littéraires jugeront contre-nature, à savoir une analyse des besoins des publics, une étude de leur horizon d'attente, exigences et attentes plus diffuses confondues, un bilan de leurs compétences techniques et culturelles. Cette analyse des besoins vise donc à déterminer pour qui et selon quel usage l'œuvre littéraire peut avoir du sens. Elle permet également de savoir comment il faut l'apprêter pour que ces nouveaux publics puissent utiliser les ressources selon leurs attentes ou pour faire naître de nouvelles idées d'utilisation.

### **Les conditions de « revitalisation » numérique du corpus**

La recherche dans ce domaine ne se pose donc pas seulement en termes de technique informatique ou de connaissances linguistiques ou littéraires. Elle implique aussi toute une réflexion sur la possibilité de mettre à disposition de nombreux utilisateurs aux objectifs et aux pratiques différents un même document de base, en l'occurrence un corpus littéraire et son iconographie annexe. Un tel travail d'équipement des corpus doit se faire en partenariat étroit entre les divers secteurs concernés : conservateurs et responsables techniques des établissements détenteurs des fonds, spécialistes des contenus linguistiques et littéraires, informaticiens et professionnels du multimédia.

Il comprendra nécessairement une phase d'équipement technique de la documentation essentiellement consacrée au marquage sémantique du corpus ainsi qu'une phase d'appareillage conceptuel plus particulièrement tournée vers les interactions et la conception de scénarios d'utilisation. Cependant, compte tenu de la spécificité du fonds évoqué, il faut d'abord s'arrêter sur la grille de lecture et d'interprétation du corpus qui sert de dénominateur commun aux divers préparateurs chargés de travailler à la préparation des « archives » en vue des usages traditionnels comme nouveaux.

### **La sémiotique du territoire comme « entrée polyvalente » dans le corpus**

Aujourd'hui, l'opération qui consiste à équiper un tel corpus associe spécialistes des contenus et de l'informatique. Elle se concentre essentiellement sur la mise au point d'appareil de notes, de renvois par des liens ancrés dans le texte vers des documents placés en annexes ou vers d'autres sites ; bref, c'est l'édition critique hypertextuelle<sup>7</sup> destinée à une diffusion et à un usage savants qui constitue le principal horizon de ces nouvelles techniques appliquées au texte littéraire. Au cours des prochaines années, compte tenu de l'élargissement des publics et des utilisations mais aussi du potentiel d'évolution technique que possède l'outil informatique, il sera indispensable de réfléchir à des formes de didactisation des documents beaucoup plus variées.

---

<sup>7</sup>Exemples : (Goulet, 2000), projets IRMA et Flaubert (cf. Bibliographie).

La préparation informatique du texte (pose de balises, outillage technique – conceptuel et ergonomique – etc.) dès sa numérisation par l’application de protocoles définis en fonction du genre des documents, devra, pour des raisons de diffusion, prendre en compte l’éventail le plus large possible des publics susceptibles de les utiliser. La didactisation de l’œuvre littéraire ou du document culturel ira donc de pair avec la mise au point d’une scénarisation des données comme des utilisations possibles.

Dans le cas de notre corpus, c’est l’objet « territoire » qui constitue à la fois le contenu sur lequel il s’agit de travailler et la forme qui va tout à la fois contraindre et faciliter la didactisation des documents. Le territoire est un *objet culturel* doté de sens documentés dans des contextes sociohistoriques donnés. Il est constitué d’un grand nombre d’informations qui vont des plus élémentaires, comme les repères géographiques, aux plus complexes qui pourraient être par exemple les « représentations » successives de ce territoire, classées selon leur apparition chronologique. Mais, le territoire est aussi un *objet interactionnel* dans la mesure où sa signification est toujours le résultat d’échanges entre divers interlocuteurs ; en somme, il est un interdiscours permanent, toujours en construction, toujours en évolution, toujours en négociation. Dans le cas d’une exploitation numérique du corpus, ces dimensions donnent toutes deux prise à la didactisation, c’est-à-dire à la préparation de la documentation et de sa transmission à des publics ciblés. Le territoire « objet culturel » donnera lieu à une opération de marquage (cf. « Prise en compte de la sémiotique du territoire par un repérage automatique ») alors que le territoire en tant « qu’objet interactionnel » sera la matière de scénarios d’interactions entre l’application informatique et les usagers (cf. « Revitalisation du corpus territorialisé par la mise à disposition d’outils »).

#### **L’appareillage technique du corpus**

##### **○ Marquage et éléments d’interprétation**

Cependant, spécialistes des contenus et informaticiens se rejoindront sur une porte d’entrée commune dans le corpus et la même grille de lecture et d’interprétation des documents. Cette dernière est fournie par la sémiotique du territoire ou plutôt par certains éléments empruntés à ce système sémiotique, à savoir les noms de lieux (toponymes, hydronymes, domonymes, etc.) référencés et documentés ainsi que les marques spatiales internes au récit et en relation avec le système énonciatif.

La sémiotique du territoire constitue donc le mode de lecture et d'interprétation commun des documents. En mettant l'accent sur les repères spatio-temporels, elle permet l'élaboration d'index sémantiques qui sont autant de points d'entrée ou d'accroche pour les différents concepteurs et préparateurs de la version informatisée du corpus. Ces repères représentent à la fois la richesse documentaire du corpus considéré et une partie essentielle de son potentiel de dynamisation. Sur le corpus considéré, nous avons donc appliqué une grille de balisage relativement fine et assez largement polyvalente puisqu'elle devrait pouvoir répondre aux besoins de plusieurs types d'utilisateurs pour des usages variés, depuis la modalité informative jusqu'à la recherche d'ordre linguistique.

Ainsi, la conception du territoire comme *contenu sociohistorique constitué* ou *objet culturel* a tout naturellement trouvé un prolongement didactique dans la démarche d'informatisation. Toutefois, si l'on considère que le territoire est un *objet culturel* complexe, il est bien évident que la documentation relative aux noms de lieux constitue une dimension minimale voire minimaliste de l'ensemble des données que cet objet culturel comprend. Cependant, comme nous le verrons plus avant, ces repères territoriaux font signe dans le texte comme sur les autres types de documents utilisables (cartes géographiques, cartes postales, photographies, et autres documents iconographiques, etc.) ou sur le terrain pour l'ensemble des préparateurs-concepteurs de l'appareillage des documents comme pour tous les utilisateurs potentiels.

- **Les interactions et les divers scénarios d'utilisation des ressources**

Associé au linguiste ou au géographe, dans un premier temps, l'informaticien tirera de ces éléments territoriaux un index qui répertoriera les segments linguistiques simples comme complexes liés à l'expression du lieu. Il les classera en fonction des précisions qui lui seront fournies. Dans un deuxième temps, cette base de données sera susceptible de permettre toutes sortes d'opérations de « manipulation » et d'appropriation des documents par des utilisateurs experts comme plus occasionnels. Le linguiste, le littéraire – et la communauté scientifique dans son ensemble – y verront la possibilité d'en faire un usage savant et exhaustif. L'enseignant en proposera un usage pédagogique à l'intention de ses élèves. Quant aux professionnels de l'information, de l'économie locale ou du tourisme, ils en feront un usage partiel et parfois inattendu – parce que détourné ou inédit – pour apprêter un service à un public restreint (restaurateurs, hôteliers, touristes, etc.) ou en direction du grand public (offices de tourisme, sites web des grandes villes ou des institutions régionales).



Ainsi, compte tenu de l'ouverture potentielle à de nouveaux publics pour de nouveaux usages, le corpus doit être préparé ou « didactisé » de manière à le rendre préhensible aux plans conceptuels comme ergonomiques dans le cadre de pratiques étendues et renouvelées qui restent aujourd'hui, pour l'essentiel, à créer tant sur le plan conceptuel que technique.

Si dans la phase de didactisation et d'appareillage des documents, les points de rencontre entre les spécialistes des contenus et leurs homologues informaticiens étaient avant tout les toponymes et autres repères spatiaux, d'autres concepts deviennent progressivement des entrées particulièrement fructueuses dès lors qu'il s'agit d'aborder les usages. Ainsi, l'approche du concept de territoire comme *interdiscours* ou *objet interactionnel* ouvre des perspectives intéressantes en termes de scénarisation du corpus notamment quant à la mise au point de scénarios d'utilisation des ressources.

Il est toutefois certain qu'en associant étroitement la connaissance du domaine local et les concepts de la sémiotique du territoire par l'intermédiaire des multiples moyens de repérage et d'accès à l'information que fournissent les technologies de l'informatique et des multimédias, ces nouveaux usages peuvent notablement modifier la relation aux corpus littéraires et culturels. Ils peuvent aussi susciter de nouvelles interprétations de leurs contenus à travers de nouvelles approches méthodologiques et un renouvellement des besoins documentaires.

#### **Contribution de l'informatique à la resocialisation d'un corpus restreint**

Dans cette partie, nous nous intéressons au point de vue informatique et technologique de la problématique posée, à savoir, si les questionnements posés ont déjà trouvé, et dans quelle mesure, une faisabilité technologique.

Notre étude démarre cependant une fois la première étape de numérisation et de reconnaissance (type Reconnaissance Optique de Caractères pour du texte) terminée. Cette étape a pour effet de transposer le corpus littéraire sur support papier en un corpus numérique, c'est-à-dire un ensemble de documents numérisés dotés d'une structure logique (paragraphe, titres) et d'une structure physique (casse, police de caractères). S'agissant d'un corpus littéraire, notons que cette transposition entraîne également pour lui, la perte de son « statut » unique d'œuvre littéraire au profit d'un double statut œuvre-document, avec ses conséquences en termes de point de vue d'analyse. Ainsi, par exemple, la notion d'analyse rhétorique de l'œuvre initiale se voit transposée en analyse de la structure logique du document résultant. Par ailleurs, le caractère restreint du corpus auquel nous nous intéressons n'a pas d'incidence particulière sur le processus de

numérisation, si ce n'est la possibilité de recours à une assistance manuelle de la phase de reconnaissance en cas de difficulté.

À partir d'un fonds numérisé, nous nous intéressons alors d'une part, à la préparation conceptuelle et technique des documents en vue d'une meilleure prise en compte de la sémiotique du territoire, et, d'autre part, à leurs modalités de revitalisation par le renouvellement de leurs usages à destination des différents publics concernés par la territorialité. Cette *démarche d'ingénierie* s'appuie nécessairement sur le travail d'une équipe interdisciplinaire où sont représentées les préoccupations et compétences des prescripteurs, bénéficiaires et metteurs en scène/préparateurs du corpus littéraire territorialisé, à savoir des spécialistes en langue et littérature, en aménagement du territoire et en informatique.

Ce travail d'ingénierie se focalise sur deux aspects. D'une part, une ingénierie du *document*, dont l'objectif est l'identification et la mise en valeur, au sein des documents, de marques de la sémiotique du territoire constituant une source d'intérêt pour les publics ciblés. D'autre part, une ingénierie des *interactions*, dont l'objectif est la formalisation des usages que les différents publics ciblés peuvent avoir des documents et la mise en œuvre des nouvelles formes d'exploitation qui en découlent. Ces deux aspects sont duels et complémentaires. En effet, et de manière analogue au domaine littéraire, et conformément à (Pédauque, 2003), le document est traité comme un « objet culturel » que l'on analyse sur le plan de sa logique interne, mais aussi comme un « objet interactionnel » que l'on reconstruit à chaque nouvelle utilisation selon des logiques externes liées à sa socialisation.

D'un point de vue technique, une première problématique s'oriente donc vers le marquage (c'est-à-dire le repérage et l'interprétation ciblée) automatique ou semi-automatique de segments d'informations, en particulier spatio-temporels, au sein de documents textuels ou cartographiques relevant d'un corpus littéraire territorialisé. Les documents numérisés deviennent alors des ressources sémantiquement marquées, interrogeables et combinables grâce aux outils d'exploitation qui les accompagnent et qui sont basés sur les balises précédemment posées. Notons que les spécificités d'un corpus littéraire territorialisé constituent un atout dans la mise en œuvre du repérage. En effet, d'une part, l'automatisation complète du processus peut être facilitée par le caractère spécifique et circonscrit des contenus analysés. D'autre part, lorsque l'automatisation n'est pas possible, c'est le caractère restreint de ce corpus qui permettra de mener à terme un marquage semi-automatique. Par ailleurs, le degré d'automatisation de ce processus varie selon le type de public auquel celui-ci est destiné. Ainsi, par exemple, un marquage destiné à un public de

spécialistes savants fera davantage appel à une supervision humaine en raison de la quantité et de la variété de marques qu'il doit produire. De même, les marques jugées d'intérêt pour un public non spécialiste seront moins nombreuses et sans doute plus faciles à identifier de manière automatisée.

La deuxième problématique s'oriente vers l'élaboration d'outils permettant l'exploitation, par les différents publics ciblés (savants, touristes, scolaires), des ressources sémantiquement marquées. Il est certain que ces outils doivent être adaptés aux objectifs d'utilisation, attentes et degrés de compétence de chaque public sur le domaine traité. Cela suppose une analyse préalable des publics ciblés et une formalisation des usages attendus. Ainsi, pour les spécialistes du domaine, les fonctionnalités développées permettront la recherche, l'indexation, l'annotation, l'hyperisation et/ou la recombinaison de nouveaux documents à partir de plusieurs sources analysées. Pour les publics scolaires ou touristes, il s'agira plutôt de scénarios prévus par les prescripteurs et permettant le parcours, l'exploration, voire l'annotation des ressources à l'intérieur d'un contexte d'interprétation familier facilitant leur lecture et leur appropriation. Les scénarios aident à formaliser les contenus des documents, mais aussi les contextes d'utilisation (usages) de ces documents ainsi que les modalités d'interaction avec les publics.

À titre d'exemple, un premier travail d'exploration permettant la validation de nos propositions a été réalisé sur un corpus textuel et cartographique issu de la littérature populaire du domaine basque, et tout particulièrement de deux textes (Cerquand, 1872), (Barbier, 1931).

Pour conclure cette partie, nous évoquons les limites de notre proposition et les perspectives.

### **Prise en compte de la sémiotique du territoire par un repérage automatique de marques**

La démarche de modélisation opératoire nécessaire au traitement automatique ou semi-automatique du contenu du type de corpus littéraire auquel nous nous intéressons part des caractéristiques spécifiques de celui-ci. En effet, en tant que fonds territorialisé, il est constitué de « segments », porteurs d'information sur le territoire et ancrés sur une localisation spatiale et temporelle. De plus, ce corpus est presque toujours composite, associant des composantes textuelles, iconographiques (en particulier des cartes), phonographiques, voire multimédias. Différentes techniques et divers outils permettent l'identification de ces différents types de marques ; les résultats les plus avancés s'inscrivent dans le marquage des composantes textuelles et cartographiques que nous présentons ci-après, tandis que les autres types de marques seront abordés dans les perspectives d'évolution de notre travail.

### Illustration des objectifs fixés à travers un exemple

Le texte suivant est un extrait d'une légende décrite dans les recherches de Cerquand (1872), et sert de support à un article publié dans *Pays Basque Magazine*, un magazine touristique et culturel destiné à faire découvrir ou mieux connaître le Pays basque au grand public (Laxalt, 2003).

*À la lisière du **bois de Zouhoure**, un pâturage est sur le versant de **la montagne Azaleguy**, et au milieu du versant, un antre dominant un abîme.*

*Autrefois les pasteurs d'alentour perdaient leur bétail et n'en trouvaient trace *nulle part*. Un jour un effroyable serpent (**Herensuge**) sortit de l'antre pour aller boire. On vit sa tête à l'eau du ruisseau, et la queue encore près de l'antre. Il attirait les brebis par sa seule aspiration et les engloutissait. Que fallait-il donc faire ?*

Source : J.-F. Cerquand, Contes parus dans *Le Bulletin des sciences, lettres et arts de Pau*, 1875-1876.

À titre d'illustration, nous avons identifié manuellement les éléments suivants :

- en gras les noms de lieux, de maison, de personne ou de personnage ancrés dans un territoire réel ;
- en italique les références spatiales ancrées dans le déroulement interne du document ;
- en souligné, les éléments à connotation temporelle.

À travers une représentation cartographique que l'on pourrait emprunter à l'IGN ou composer à partir d'un autre type de fond de carte, il serait possible de proposer au lecteur-touriste un parcours initiatique permettant de découvrir la région dans laquelle se situe le conte. Ceci pourrait aussi être complété par des images et chants circonstanciés.

Il s'agit donc de proposer des techniques et outils informatiques permettant d'effectuer tout ou partie de ce marquage et d'assister l'expert dans cette tâche. Bien que les éléments identifiés dans l'exemple soient tous de même nature spatio-temporelle, nous verrons que les outils et techniques permettant leur identification automatique ne sont pas de même niveau de complexité. En effet, un premier ensemble d'outils permet de repérer automatiquement les termes en gras représentant des toponymes et domonymes, tels que « Zouhoure » ou « Azaleguy ». Ce repérage serait facilité par exemple par l'utilisation d'un atlas géographique et linguistique (Videgain, à paraître). Le processus de traitement nécessite aussi un autre ensemble d'outils permettant alors d'identifier, par extension, des termes ou expressions tels que « du bois », « à la lisière » ou « de la montagne », ou encore d'interpréter des segments plus complexes tels que « À la lisière du bois de Zouhoure ».

D'autres types de traitements pourraient également compléter cette analyse en identifiant de nouvelles balises telles que :

- des images relatives aux lieux décrits dans les textes, dont on pourrait extraire des éléments significatifs ;
- un ensemble d'iconographies relatives à la période des documents textuels (si des éléments d'époque sont significatifs) ;
- des objets sonores ou vidéos.

### **Analyse spatio-temporelle textuelle**

#### **o Principes**

Un premier traitement sur le texte permet le repérage des termes candidats grâce à une analyse typographique (ici, tous les mots qui commencent par une majuscule) et morphologique (ici, tous les noms, car un *nom* se définit comme un mot ou un groupe de mots qui nomme une personne, un lieu, une chose ou un concept).

Puis, les termes ainsi marqués peuvent alors être soumis comme requête à un *service* permettant de le situer par exemple dans un découpage administratif. Ce service peut être un site web, par exemple comme celui du Quid <<http://www.quid.fr>>, qui, interrogé avec l'exemple du paragraphe précédent, nous a répondu en associant « Zouhoure » à la commune d'Alçay-Alçabéhéty-Sunharette.

Dans un troisième temps, cette nouvelle information nous permet de questionner un autre service permettant une localisation géoréférencée directe ou par le biais du découpage administratif englobant. Pour notre exemple, nous avons utilisé le service disponible sur le site du Conseil Général des Pyrénées Atlantiques <[http://www.cg64.fr/fr/atlas\\_departement.asp](http://www.cg64.fr/fr/atlas_departement.asp)>, et avons émis la requête à partir du nom de la commune « Alçay ». Ce traitement nous permet déjà de construire une interprétation et d'adopter l'hypothèse que l'entité nommée est une entité à connotation spatiale.

Enfin, il s'agit d'appliquer une analyse morfo-syntaxique sur l'ensemble des termes environnant l'entité marquée et référencée spatialement. Ce traitement, limité à une certaine profondeur (la phrase contenant l'entité nommée), permet de marquer les différents mots qui sont éventuellement associés à cette entité. Les nouveaux termes repérés vont permettre de réaliser une interprétation spatiale plus fine de l'entité nommée. Dans l'exemple, deux nouveaux noms, « bois » et « lisière » sont ainsi marqués.

À ce niveau, pour une interprétation plus fine du syntagme nominal, « À la lisière du bois de Zouhoure », les différentes ressources nécessaires sont de différents types : il peut s'agir de lexiques tels que des *dictionnaires* des synonymes <<http://elsap1.unicaen.fr/dicosyn.html>>, ou encore

d'ontologies modélisant les principaux concepts de positionnement statique (« au bord », « dans », « entre », etc.) ou d'orientation (« Nord », « Sud », « devant », « à droite de », etc.).

L'analyseur temporel fonctionne selon les mêmes principes. Les « points d'arrêt » sont alors des dates, des opérateurs définissant des intervalles (« de X à Y », « entre X et Y », « les années X ») et ils sont soumis à une nouvelle classe d'opérateurs, tels que (« le début de X », « aux alentours de X », etc.).

Les différents principes exposés montrent que la démarche adoptée permet, d'une part, de régler le niveau d'automatisation du processus et, d'autre part, de graduer la finesse des résultats attendus. Nous allons ci-après présenter les outils techniques sous-jacents permettant de mettre en œuvre les principes évoqués.

#### ○ Outils

Les travaux que nous citons sont consacrés à l'aide à l'analyse de documents territorialisés, notamment dans leurs composantes textuelles et cartographiques. Ces outils d'analyse calculent des représentations sémantiques qui seront ensuite exploitables dans les procédures de recherche de segments documentaires.

Le cœur de la méthode consiste à :

- repérer et analyser, dans les composantes textuelles et/ou cartographiques, des « segments » porteurs d'information territorialisée (c'est-à-dire spatiale et temporelle) ;
- donner à ces segments des représentations symboliques indépendants du mode d'expression ;
- associer ces représentations symboliques aux occurrences des segments repérés à l'intérieur du document.

Cette méthode aboutit à une indexation intra-textuelle, ou plus généralement intra-documentaire, du document analysé. Nous proposons d'utiliser un marquage XML (« *eXtended Markup Language* ») des différents termes identifiés. Ceux-ci seront aussi mis en relations et exploitables par un explorateur XML, c'est-à-dire indépendamment de la présentation des contenus.

Ainsi, une entité nommée à connotation spatiale ou temporelle sera interprétée après repérage dans le texte – ou dans une carte – selon trois différents points de vue :

- granularité (niveau de précision : région, commune, bois, etc.) ;
- position (lisière, dans, entre, etc.) ;
- orientation (sud, nord, devant, etc.).

Ces éléments sont formalisés sous la forme d'attribut/valeur, ce qui permet également de construire la représentation sémantique de l'entité en XML (Gaio, 2001).

De plus, notre approche du traitement de l'information territorialisée combine des procédés automatisés et semi-automatisés, tant pour l'analyse de composantes textuelles que cartographiques. Elle utilise la technique des grammaires sémantiques locales, implémentable dans Intex <<http://intex.univ-fcomte.fr/>>, et parfaitement adaptée ici puisqu'il s'agit d'analyser des syntagmes nominaux ou prépositionnels *contigus*. Grâce au modèle de marquage XML adopté, l'ensemble des outils nécessaires à la mise en œuvre de la démarche est intégrable dans la plate-forme LinguaStream <<http://www.linguastream.org/>>, développée pour répondre à une problématique très proche sur le plan de l'analyse textuelle. Ceci permet donc un gain de temps considérable dans l'implémentation d'un prototype du système de marquage sémantique.

Nous disposons ainsi d'une chaîne de traitements combinant les différents analyseurs relatifs aux différents aspects énoncés dans la section précédente (cf. « Illustration des objectifs fixés à travers un exemple »). Conformément aux besoins exprimés, le système permet de gérer un nombre quelconque de « couches de balisage », qui se superposent à la structure logique du document.

#### **Analyse des composantes cartographiques**

Les éléments non textuels d'un corpus territorialisé sont multiples : documents cartographiques, iconographiques (structurés ou non), sonores, vidéos, composites, etc. Or, leur étude et utilisation en extraction d'information posent un certain nombre de problèmes en raison du caractère spécifique de chaque médium (cartes, dessins, extraits sonores, extraits vidéos). Divers travaux de recherche existent dans le domaine de l'analyse de contenus non textuels, avec des degrés d'aboutissement variables. Parmi eux, les travaux effectués dans le domaine cartographique méritent d'être précisés en raison des résultats obtenus. Il est toutefois intéressant de noter que, lorsque cela est possible, tous les travaux s'appuient sur les méthodologies stables et éprouvées du domaine textuel.

Les documents cartographiques constituent un média largement employé pour communiquer de l'information spatialisée. Dans le cadre d'un traitement automatique du contenu, le problème est de définir un modèle opératoire pour ce mode d'expression afin de pouvoir élaborer des traitements équivalents à ceux disponibles pour le texte. Nombreux travaux (Carrière, 2000 ; Pratt, 1993 ; Head, 1984 ; Bertin, 1973) permettent de considérer l'expression cartographique comme un mode d'expression à part entière au même titre que la langue naturelle. Ceci nous conduit à regarder la



carte du point de vue de la sémiotique. En effet, l'expression cartographique peut être vue comme un système de communication qui consiste à utiliser des signes dans un système de signification : la carte. De manière plus précise, un signe est constitué de deux composants : sa facette physique, qui dépend de la forme de communication (ici une marque sur un papier ou sur un écran) et habituellement appelée l'*expression* du signe, et sa facette conceptuelle, qui réfère à une signification que nous pouvons appeler son *contenu*. La carte, système de signification, doit être liée à un *code* commun et connu par l'ensemble des acteurs impliqués dans le processus de communication. Cela nous amène donc à la délicate question de la sémantique des cartes, objet de plusieurs travaux dans le domaine de la cognition, comme attesté par T. Barkowsky et C. Freska (Barkowsky et Freska, 1997). J. Bertin (*op. cit*) a montré que la sémantique d'une carte émerge nécessairement de l'explicitation des signes utilisés ainsi que de leur structuration syntaxique. Mais, étant donné que les deux composantes du terme sémiotique « signe » sont, d'un côté, les entités cartographiques placées sur la carte et, de l'autre côté, la connaissance géographique que nous leur faisons correspondre, cette syntaxe est en partie dépendante de la tâche de lecture considérée. Autrement dit, lors de la conception d'un traitement automatique du contenu des cartes, il faut *a priori* être capable de définir, selon la tâche de lecture attendue, quelles sont les entités contenues dans les cartes qui expriment le contenu, et quelles sont celles qui représentent le contexte, ceci afin de dégager le code qui les lie.

Par exemple, dans le cas d'une tâche de lecture se limitant à un repérage des entités nommées, les segments d'information à repérer seront constitués des toponymes et de leurs marques visuelles de localisation (point, symbole, surface, etc.). Alors que dans le cas d'une tâche de lecture consistant à repérer l'organisation spatiale d'un fait en termes d'uniformité et de contraste, les seules entités porteuses de contenu seront des entités exclusivement constituées par des objets de forme surfacique dont l'expression permet d'observer leur degré de similarité.

Nous avons montré l'intérêt du marquage des documents territorialisés et nous allons maintenant montrer de quelle manière il peut être exploité par des outils appropriés en vue de faciliter la resocialisation du corpus.

### **Revitalisation du corpus territorialisé par la mise à disposition d'outils d'exploitation des ressources numérisées et sémantiquement marquées**

La resocialisation d'un corpus numérisé passe par sa revitalisation numérique, c'est-à-dire par sa mise à disposition de nouveaux publics pour de nouveaux usages. D'un point de vue technologique, la problématique consiste donc à concevoir les outils permettant l'exploitation de

ces ressources sémantiquement marquées par les différents publics ciblés, et, en amont de la conception, elle consiste à analyser ces publics en termes de compétences, d'attentes, et à spécifier quels sont les usages attendus du corpus (cf. « Nouveaux publics, nouveaux usages »).

Un premier travail d'exploration nous a permis d'identifier trois catégories de publics (scientifique, touristique, scolaire) auxquelles sont associées des usages particuliers du corpus. Cette analyse a été présentée dans la première partie de cet article. Elle constitue une base pour la spécification des outils d'exploitation adaptés à chaque type de public, en termes d'éléments de description fonctionnels et interactionnels. Les descripteurs fonctionnels aident à la spécification fonctionnelle des services attendus, les descripteurs interactionnels aident à la spécification des modalités et/ou conditions d'utilisation de ces services par les publics ciblés, en fonction de leurs compétences et horizons d'attente.

Afin d'illustrer en partie le travail restant à faire sur la spécification des outils, nous présentons, dans le paragraphe suivant, quelques éléments de description fonctionnelle et interactionnelle, en sélectionnant deux cas de figure, d'une part, les outils d'exploitation destinés à la communauté scientifique, et d'autre part, les outils destinés à une communauté grand public.

#### **Outils d'analyse, de recherche, d'annotation et de recomposition des ressources marquées**

Nous nous intéressons ici aux outils permettant, à un utilisateur expert du domaine (savant, chercheur), de prolonger, voire d'enrichir, ses activités de recherche traditionnelles, à savoir, une étude approfondie du texte sur un corpus numérisé.

Ces activités se basent sur l'analyse des œuvres, selon de multiples critères (comme les lieux, les personnages, les faits, les styles, la construction du texte – syntagmatique, rhétorique, etc.). Les analyses donnent lieu à la production d'éditions savantes, c'est-à-dire des éditions annotées des œuvres analysées.

Ainsi, d'un point de vue fonctionnel, les services développés doivent faciliter la recherche, l'indexation, l'hyperisation, l'annotation et la recomposition de nouveaux documents à partir de plusieurs sources analysées.

La numérisation fournit de fait d'intéressants outils de recherche, d'analyse et d'indexation de type documentaire, basés pour la plupart d'entre eux, sur les méta-informations associées aux documents, sur leur structure logique et, de manière moins riche, sur leur contenu. De plus, la conservation de documents dans un format XML facilite la mise à disposition d'outils d'annotation

et d'hypérisation permettant la mise en œuvre d'éditions critiques. Toutefois, ces fonctionnalités ne sont pas suffisamment riches pour un public expert, elles doivent être complétées par des outils de recherche et d'analyse également basés sur les contenus sémantiques du corpus. Dans cette perspective, les outils d'exploitation d'un marquage spatio-temporel sont un apport fondamental à l'acceptation de l'outil informatique comme support au travail des experts linguistiques et littéraires.

Le marquage spatio-temporel du corpus fournit un ensemble de documents munis d'une indexation multidimensionnelle, c'est-à-dire de plusieurs « couches de balisage » qui se superposent à la structure logique du document. Ainsi, une exploitation des documents marqués à l'aide d'un explorateur XML enrichit d'ores et déjà considérablement les possibilités de recherche documentaire disponibles. Par ailleurs, les outils de marquage précédemment décrits (cf. « Analyse spatio-temporelle textuelle ») peuvent également servir aux experts pour de nouveaux marquages ou de nouvelles analyses, par exemple, à l'occasion de la publication, par la communauté scientifique, de nouvelles ressources linguistiques et/ou géographiques extérieures permettant la validation et/ou l'ancrage géographique de ces nouvelles analyses spatio-temporelles.

D'un point de vue interactionnel, on peut constater que les services identifiés combinent des procédés automatisés et semi-automatisés et ne relèvent pas d'un même niveau de complexité d'utilisation. Le public, expert en littérature et linguistique, ne maîtrise pas forcément la recherche documentaire, ni les divers outils d'analyse, de marquage cités. Ainsi, toute mise en exploitation informatique doit prévoir l'intégration de ces services à l'intérieur de plates-formes comptant un serveur d'application, des outils de navigation, de visualisation, d'édition hypertextuelle, des interfaces spécifiques, des outils d'assistance adaptés à divers niveaux d'expertise technologique des utilisateurs, et pouvant aller jusqu'à la proposition de scénarios d'utilisation préétablis.

#### **Environnements scénarisés de parcours, découverte et consultation des ressources marquées**

Concernant les outils permettant à un utilisateur non expert du domaine (scolaire, grand public) d'accéder à un corpus numérisé, l'analyse nous oriente plutôt vers des outils d'accès facilement disponibles dans des environnements numériques familiers (musées virtuels, bornes interactives, classes patrimoine, sites web, etc.) et associés à des activités coutumières de ces publics.

Ainsi, d'un point de vue fonctionnel, les services développés, essentiellement basés sur la navigation, doivent faciliter, voire orienter, l'exploration et la découverte d'un extrait du corpus en

immersion dans un ensemble plus vaste de ressources extérieures (géographie, hydrologie, histoire, faune, flore, vieilles cartes postales, archives sonores, hôtellerie, restauration, artisanat, circuits pédestres, équestres, etc.) qui replacent la découverte dans une activité plus large, comme par exemple, un parcours touristique, ou bien pédagogique. Des services de recherche et d'annotation doivent également accompagner la navigation pour permettre à chaque utilisateur de personnaliser et de s'appropriier les parcours effectués.

Le marquage spatio-temporel disponible sur le corpus serait alors exploité par le système de navigation, de manière transparente pour le public. En effet, les balises (noms de lieux, de maisons, positions, dates, époques, etc.) constituent autant de possibles points de connexion directs entre l'ensemble des ressources du parcours découverte (touristique, pédagogique) et le corpus qu'il est possible d'exploiter sous la forme d'options de navigation, complémentaires ou alternatives. De plus, à partir des ressources du parcours et grâce aux outils d'analyse sémantique du corpus, de nouveaux points de connexion peuvent être calculés dans le corpus, afin d'enrichir encore les possibilités de navigation. Le marquage spatio-temporel et les outils d'analyse sémantique sont donc exploités, non pas directement par l'utilisateur, mais par le service de présentation des informations, qui propose des parcours organisés d'un ensemble de ressources, en forte interaction avec le corpus auquel elles sont associées.

Ainsi, le succès de ces services dépendra davantage de leurs qualités interactionnelles, de leur facilité d'appropriation et de leur adéquation avec les attentes des publics ciblés que de l'exhaustivité ou de la finesse d'exploitation des marques situées à l'intérieur du corpus.

D'un point de vue interactionnel, ces services peuvent alors être décrits comme des environnements scénarisés de parcours de ressources marquées. Nous entendons par « scénario » un ensemble de séquences d'interactions obéissant à des conventions de communication identifiées par un « genre », c'est-à-dire plaçant l'utilisateur dans un contexte d'utilisation et d'interprétation qui lui sont familiers. Parmi les genres d'applications identifiés, nous pouvons citer les applications ludiques, éducatives, encyclopédiques, informatives, promotionnelles, etc. Ainsi colorisés d'un ou plusieurs genres, les scénarios permettent une transmission organisée des contenus de l'application, d'une part, en accord avec les intentions de communication souhaitées par les prescripteurs et, d'autre part, inscrite dans un contexte d'utilisation familier de l'utilisateur (Latapy, 2004).

À titre d'illustration, les ressources et techniques dont nous disposons nous permettent d'évoquer le scénario d'un parcours découverte de type documentaire, à usage touristique. Ainsi,

par exemple, dans un ensemble de ressources géographiques locales qui lui sont proposées, l'utilisateur indique, le point de départ de la visite qu'il souhaite effectuer, le point d'arrivée et, éventuellement, les étapes qu'il compte faire. Le système calcule alors, dans un premier temps, un premier parcours composé des points choisis accompagnés de différentes ressources extérieures disponibles (informations climatiques, historiques, gastronomiques, hébergement, produits régionaux, curiosités locales, etc.). Dans un deuxième temps, grâce aux balises et aux outils de recherche spatio-temporelle disponibles, il puise dans le corpus littéraire balisé, des textes ou extraits de textes littéraires illustrant les étapes précédemment sélectionnées. Puis, grâce aux outils d'analyse spatio-temporelle disponibles, il puise, également dans le corpus, d'autres lieux géographiquement proches de ces étapes. Ainsi, les nouveaux lieux issus du corpus permettent de compléter et d'enrichir le parcours initialement calculé à partir des seules ressources extérieures, alors que les textes permettent d'enrichir les ressources extérieures initialement sélectionnées. Pour finir, c'est ce parcours doublement enrichi qui est proposé au visiteur.

### **Limites de nos propositions, perspectives d'évolution**

Nous avons montré comment un processus de numérisation, basé sur une exploitation de connaissances spatio-temporelles et d'index sémantiques, permet une réelle revitalisation d'un corpus littéraire restreint. Le résultat visé est un corpus didactisé et mis en valeur par des techniques de scénarisation. Les outils et techniques que nous proposons entraînent un changement des modalités de lecture, d'appropriation, de maniabilité technique et, *in fine*, d'interprétation du document. La didactisation rejoint la définition étymologique grecque de transmission du savoir dans la mesure où nous proposons un environnement accompagnant les divers utilisateurs humains dans les usages qu'ils se définissent.

Notre proposition se nourrit de divers travaux réalisés dans des contextes différents de celui du corpus littéraire et aboutissant à des résultats avérés. Toutefois, elle connaît sur le corpus un certain nombre de limites qui constituent pour nous autant de pistes pour des travaux futurs.

Une première limite concerne la part restreinte du corpus territorialisé pouvant être traitée par les outils de marquage que nous proposons.

En effet, d'une part, ces outils ne prennent en compte que des ressources textuelles et cartographiques. Or, les fonds territoriaux sont constitués de documents extrêmement variés (cartes postales, photographies, extraits sonores ou vidéos, dessins, etc.) qu'il est impossible d'écarter, du moins à long terme, d'un projet de numérisation et de revitalisation du patrimoine. Il en résulte, d'un point de vue technologique, la nécessité d'élargir le champ d'action des outils de marquage et d'analyse proposés, notamment à des composantes iconographiques et multimédias. D'autre part,

ces outils ne permettent pas une analyse aussi fine pour tous les types de textes disponibles, comme par exemple les textes fictionnels. En effet, notre méthodologie, qui ne peut trouver que des marques typo-dispositionnelles et syntaxiques, et non spatiales, s'appuie sur des ressources extérieures (atlas géographiques, etc.) qui n'existent pas dans le cas de textes fictionnels.

Concernant la prise en compte de documents iconographiques (photographies, dessins, illustrations), des techniques d'analyse sont aujourd'hui opérationnelles sous la forme de produits commerciaux. On peut notamment citer le projet Imedia (Boujemaa et Jurie, 2003), basé sur la segmentation des images en différents objets identifiés de manière unique selon une série de descripteurs (couleur, texture, forme, configuration spatiale...). Les identifiants sont ensuite analysés relativement à une base de connaissances. De plus, lorsque des composantes textuelles, telles que titres et légendes, accompagnent une iconographie, une chaîne de traitement contenant des analyseurs textuels peut leur être appliqué, permettant alors d'obtenir un marquage compatible avec les informations extraites des composantes textuelles.

Concernant la prise en compte de documents sonores, les techniques d'analyse existantes varient selon qu'ils contiennent ou non de la parole. Dans le cas où la parole est présente, l'analyse se fait généralement en deux étapes (Garofolo, *et alii*, 2000). La première étape consiste à transcrire, à l'aide d'un système de reconnaissance automatique de la parole, les documents audios comprenant de la parole en du texte enrichi de méta-données, afin de conserver toutes les informations (temps de parole, changement de locuteur, etc.). La deuxième étape consiste à appliquer des techniques de traitement automatique des langues sur le texte enrichi, en prenant en compte les méta-données ainsi que la mesure de confiance affectée à chaque mot lors de la transcription automatique (erreurs de transcriptions, absence de ponctuations, etc.). Ainsi, dans ce cas, le traitement du son se rapproche du traitement du texte après reconnaissance vocale (Vinet, 2004) avec la terminologie MPEG-7 (Martinez, *et alii*, 2002).

Concernant la prise en compte de documents vidéos (Lefèvre, 2002), (Palakal et Tennant, 2002), les systèmes d'analyse existants se basent sur deux principaux concepts : la segmentation temporelle, permettant une détection des changements de plan, et l'analyse des contenus pour chacun des plans préalablement identifiés. La détection des changements de plan est basée sur une analyse globale des couleurs, de la texture et des mouvements. L'analyse des contenus a pour objectif d'identifier les objets pertinents, c'est-à-dire ne faisant pas partie du fond de l'image. Les techniques d'analyse tirent également partie des diverses composantes incluses dans un document vidéo : dialogues, sous-titrages, bande sonore. Ainsi, l'analyse de l'image animée est complétée par les résultats des analyses des différentes composantes.

Une deuxième limite concerne la détermination de la portée des références spatio-temporelles identifiées. En effet, lorsque l'analyse détecte une marque spatio-temporelle à l'intérieur d'une phrase, le problème qui se pose alors est de savoir délimiter la portée de cette marque, c'est-à-dire, la portion de texte (phrase, paragraphe, groupe de paragraphes) se rapportant sémantiquement à cette marque. C'est la connaissance du couple (marque, portée de cette marque) qui permettra une exploitation automatique ultérieure de l'extrait de texte contenant la marque identifiée, comme dans l'exemple cité à la fin de la section précédente (cf. « Environnements scénarisés de parcours »). Dans ce cas, nous pensons que d'autres techniques de marquage peuvent apporter une aide à la résolution de la problématique posée, comme la théorie de l'encadrement du discours, ou encore la théorie des structures récursives.

Dans la *théorie de l'encadrement du discours* (Charolles, 1997), l'auteur identifie des segments – *cadres de discours* – homogènes par rapport à un critère sémantique (par exemple une localisation spatiale/temporelle). Ce critère est spécifié par une expression détachée en initiale de phrase (donc aisément repérable par un outil automatique) appelée *introduceur de cadre*. Ainsi, cette segmentation permet de « *subdiviser et de répartir les informations* apportées par le discours au fur et à mesure de son développement ». À titre d'exemple, le terme « Autrefois » est un cadre de discours temporel figurant dans l'extrait de texte présenté dans la section intitulée « Illustration des objectifs fixés à travers un exemple ». Une limite de cette théorie concerne néanmoins la difficulté à savoir fermer les cadres ouverts.

Quant à la théorie des structures récursives (RST – *Rhetorical Structure Theory*) (Mann, 1999), elle avait été initialement conçue dans le cadre des études sur la génération automatique de texte pour proposer une explication de la cohérence des textes en identifiant des unités et une représentation graphique pour les structures de texte. La RST fournit une analyse pour tout texte élaboré et cohérent, et cette analyse donne une explication argumentée de la raison pour laquelle chaque élément du texte y a été inclus par l'auteur. En postulant l'existence d'une structure reliant les phrases entre elles, elle rend compte de la cohérence textuelle, d'une façon indépendante des formes lexicales et grammaticales du texte. Ce type de structures discursives apparaît immédiatement d'une grande pertinence par rapport au corpus et à la tâche visée. En effet, tout d'abord, les éléments de cohérence fournis peuvent apporter une aide à l'estimation de la portée d'une marque spatio-temporelle précédemment identifiée. De plus, l'aspect relatif à la production de cohérence des textes peut également être directement utilisé par un public de type savant souhaitant réaliser une lecture critique, ou bien apporter des éléments à une interprétation communément admise.



Ainsi, les efforts devront aller dans le sens de la mise en cohérence et de l'usage conjoint de ces différentes techniques. Notons cependant que la mise en cohérence de ces résultats avec nos travaux est une réelle difficulté. En effet, d'une part, les concepts traités et les niveaux d'abstraction abordés ne sont pas les mêmes. Ainsi, par exemple, les indices de marquage pour les composants iconographiques et multimédias, comme la texture, la couleur, etc. sont d'un niveau d'abstraction plus bas que les marques sémantiques relevées au sein d'un texte, pour lesquelles une nature spatiale ou temporelle est « aisément » repérable avec des outils existants. D'autre part, tous les travaux ne relèvent pas du même niveau d'automatisation. C'est le cas, en effet, des croquis et images, plus difficiles à traiter, qui nécessitent un marquage manuel en amont de la part des experts. Ce travail d'interprétation et d'homogénéisation s'avère encore aujourd'hui difficile à automatiser.

## Conclusion

On le voit, la numérisation représente un tournant dans le cycle de vie du document. Du reste, il serait plus juste d'employer l'expression « cycle des vies » car c'est bien une nouvelle vie sous une nouvelle forme et selon des modalités d'utilisation et de socialisation élargies et amplement renouvelées qui se profile pour le corpus « oublié » que nous avons évoqué.

On peut penser que la numérisation du document « papier » recèle sur le plan technique un véritable potentiel de revitalisation pour les contenus convertis. Cependant, chaque corpus possède sa spécificité et le projet de numérisation sera lié aux perspectives ultérieures de valorisation de ce patrimoine. Les conditions de sa resocialisation seront nécessairement tributaires des points forts qui se dégagent d'une relecture actualisée et historicisée de l'archive.

Dans le cadre de notre intervention, c'est la problématique de la revitalisation des corpus restreints qui a été envisagée et plus particulièrement celle du patrimoine local ou territorialisé. Nous avons circonscrit un corpus « oublié » et nous l'avons revisité en nous appuyant sur une grille élaborée à partir de la sémiotique du territoire qui a constitué le point d'entrée commun aux chercheurs issus d'horizon et de spécialités aussi différentes que la littérature et l'informatique. En nous appuyant sur la dimension « d'objet culturel » du territoire nous avons obtenu un ensemble de marqueurs territoriaux utiles pour de nombreux usagers. De même, la dimension « interactionnelle » du territoire a permis de proposer l'ébauche de scénarios d'interactions susceptibles de diversifier les usages traditionnels du corpus. Ainsi, nous avons pu envisager trois types d'usages : savant, pédagogique, et touristique, qui renouvellent l'approche du corpus converti en documentation électronique. C'est donc une indéniable évolution qui se produit à cette occasion

puisque les modalités de lecture, d'appropriation et, par voie de conséquence, d'interprétation du corpus, sont largement modifiées.

Toutefois, les outils de traitement de la documentation territorialisée proposés ici n'ont pour l'instant été expérimentés que sur des textes qui s'inscrivaient au sein d'un territoire et de lieux explicitement référencés dans d'autres types de documents, à savoir des cartes géographiques ou des supports iconographiques (cartes postales, photographies, etc.). Cet ancrage référentiel constitue l'une des conditions de la resocialisation du corpus territorialisé. Il reste à aborder le redoutable problème posé par le texte de fiction. Cependant, les résultats obtenus laissent espérer qu'une transposition des procédures de marquage comme d'exploitation des documents à dimension référentielle est en partie possible à la condition de créer les ressources externes adaptées à l'univers abordé. Il faut notamment évoquer l'élaboration d'un panorama complet du système autoréférentiel élaboré dans le cadre du texte de fiction ou d'un corpus fictionnel consacré à un territoire<sup>8</sup> afin de traiter la dimension spatiale de cet univers imaginaire comme un espace tangible.

Désormais, la numérisation se conçoit de plus en plus autour de projets de diffusion et de valorisation du patrimoine auprès de publics spécialisés (communauté éducative) ou plus ouverts (tourisme). Dans le cas des corpus d'intérêt local à l'inverse des grands fonds bibliographiques reconnus et plébiscités par les institutions scolaire et éditoriale et consacrés par la communauté universitaire, le critère scientifique n'est peut-être pas suffisant. C'est donc bien un projet culturel d'ensemble dans le cadre plus vaste d'un projet d'aménagement du territoire qu'il faut concevoir pour envisager de façon réaliste l'avenir des fonds numérisés et leur transmission. En effet, la mise en valeur numérique des ressources est liée à la création de lieux d'exposition (musées, écomusées, salles publiques, offices de tourisme, lieux privés, etc.) ou de pratique (centres scolaires d'initiation au patrimoine, etc.), réels ou virtuels (sites, bornes interactives, etc.) et cela demande l'implication croissante des instances et des institutions les plus diverses en tant qu'initiateurs ou partenaires des projets.

La conversion numérique des corpus bibliographiques ou culturels n'est donc pas une procédure fermée sur elle-même. À cette opération technique initiale sont liées la question de la diffusion des fonds numérisés et la problématique complexe de leur communication. Elle ouvre de nouveaux horizons en matière de publics et d'usages et invite les spécialistes des disciplines

---

<sup>8</sup> L'exemple du *Seigneur des anneaux* de Tolkien (2003) vient immédiatement à l'esprit.

concernées à renouveler leur rapport aux documents qu'ils étudient dans le silence des bibliothèques.

## Bibliographie

BARBIER, J., *Légendes basques*, Paris, Éditions Delagrave, 1931.

BARKOWSKY, T., FRESKA, C., « Cognitive requirements on making and interpreting maps », dans *Spatial Information theory : A Theoretical Basis for GIS*, Hirtle S. & Franck A. (coord.), International Conference COSIT, Laurel Highlands (USA), Éditions Springer, 1997, p. 347-361.

BERTIN, J., *Sémiologie graphique*, 2<sup>e</sup> éd., Paris, La Haye, Éditions Mouton & Cie, 1973.

BOUJEMAA, N., JURIE, F., « Recherche d'information par le contenu visuel », numéro spécial de *Technique et science informatiques RSTI série TSI*, Vol. 22, n°9, 2003, 156 p. <<http://www-rocq.inria.fr/imedia/index.html>>

CARRIÈRE, J., « Le paradigme de la communication cartographique : la hiérarchisation du *moi* et du *ici* », *Les ancrages du corps propre*, Montréal, éditions XYZ, sept. 2000, p. 87-110.

CERQUAND, J.-F., « Le chevalier de Çaro et le dragon », *Légendes et récits populaires du Pays Basque*, récité par M. Etchebarne, transcrit par M. Basterreix, dialecte souletin, orthographe de la langue basque conforme à l'édition originale des traductions parue en 1872 dans le *Bulletin des sciences, lettres et arts de Pau*, p. 211 à 213, 1872.

CHAROLLES, M., « L'encadrement du discours - Univers, champs, domaines et espace », *Cahier de recherche linguistique*, n°6, 1997.

FRÉMONT, A., *La Région, espace vécu*, Paris, Flammarion, 1976-1999.

GAIO, M., *Traitements de l'information géographique : représentations et structures*, Habilitation à Diriger les Recherches, Université de Caen Basse Normandie, décembre 2001.

GAROFOLO, J.-S., AUZANNE, C., VOORHEES, E.-M., « The TREC Spoken Document Retrieval Track: A Success Story », *Proceedings of the TREC Workshop*, Avril 2000.

GOULET, A., *Édition génétique des Caves du Vatican d'André Gide*, Université de Sheffield, CD-Rom, Paris, Gallimard, 2000.

GRANT, C., Head, « The Map as Natural Language : A Paradigm for Understanding », *New Insights in Cartographic Communication*, Christopher Board (coord.), Monograph 31, Cartographica 21, n°1, 1984.

LATAPY, M., LOPISTÉGUY, P., DAGORRET, P., « A genre approach to facilitate the intention of an application. A pedagogic experiment: risks prevention on the beach », *EISTA'04 International Conference on Education and Information Systems: Technologies and Applications*, Orlando, Florida, USA, 21-25 juillet 2004.

LAXALT, T., « Forêt des Arbailles - le souffle de la légende », *Pays Basque Magazine*, n 32, 4<sup>e</sup> trimestre, 2003, p. 40 à 45.

LEFÈVRE, S., *Détection d'événements dans une séquence vidéo*, Thèse de doctorat de l'université de Tours, décembre 2002.

LOTI, P., *Ramuntcho*, Paris, 1897, Éditions Presses-Pocket, 1986.

MANN, W.-C., *Rhetorical Structure Theory (RTS) web site*, <<http://www.sfu.ca/rst/>> [février 2007].

MARTINEZ, J.-M., KOENEN, R., PEREIRA, F., « MPEG-7, The Generic Multimedia Content Description Standard », *IEEE Multimedia*, avril-juin 2002, p.78-88.

PALAKAL, M., TENNANT, S., *CLIOH: Cultural Digital Library Indexing Our Heritage*, Indiana University Forum on Digital Libraries, <<http://clioh.informatics.iupui.edu/>> [février 2007]; <<http://www.dlib.indiana.edu/dlforum/>> [février 2007], April 16, 2002.

PÉDAUQUE, R.-T., « Document : forme, signe et médium, les re-formulations du numérique », *Working paper, Archive Ouverte en Sciences de l'Information et de la Communication*, 8 juillet 2003, <[http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/11/index\\_fr.htm](http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/11/index_fr.htm)> [février 2007].

VIDEGAIN, C., *et alii, Atlas linguistique de la langue basque*, sous l'égide l'Académie de la Langue Basque, à paraître.

VINET, H., *Description des contenus musicaux et applications*, Conférence CIDE'2004, La Rochelle, 2004.

### **Sciences de l'information et de la communication**

BERNARD, M., *Introduction aux études littéraires assistées par ordinateur*, Paris, PUF, 1999.

BRUNET, R., *Le Déchiffrement du Monde. Théorie et pratique de la géographie*, Paris, Belin, 2001, collection Mappemonde.

CHARTIER, R., *Le Livre en révolution*, Paris, Textuel, 1997.

DEBRAY, R., *Introduction à la médiologie*, Paris, PUF, 2000.

FERRAND, N., *Banque de données et hypertextes pour l'étude du roman*, Paris, PUF, 1997.

HOOG, E., « Tout garder ? Les dilemmes de la mémoire à l'âge médiatique », *Le Débat*, n°125, 2003.

JEANNERET, Y., *Y a-t-il vraiment des technologies de l'information ?*, Septentrion, 2000.

JOLY, M., *Introduction à l'analyse de l'image*, Paris, Nathan, 1993.

KLINKERBERG, J.-M., *Précis de sémiotique générale*, Bruxelles, De Boeck, 1996.

LÉVY, P., *Les Technologies de l'intelligence*, Paris, La Découverte, 1990.

LANCIEN, T., (coord.), *Multimédia : les mutations du texte*, Paris, Éditions ENS, 2000.

TOLKIEN, J.R.R., *Le Seigneur des anneaux*, Paris, Christian Bourgois, 2003.

VANDENDORPE, Ch., *Du papyrus à l'hypertexte*, Paris, La Découverte, 1999.

### **Géographie et anthropologie**

CLAVAL, P., *La Géographie culturelle*, Paris, Nathan, 1995.

De CERTEAU, M., *L'Écriture de l'histoire*, Paris, Gallimard, 1975.

Di MÉO, G., *Géographie sociale et territoires*, Paris, Nathan, 1998.

RAULIN, A., *Anthropologie urbaine*, Paris, Armand Colin, 2001.

ZONABEND, F., *Mœurs normandes*, Paris, Christian Bourgois, 2003.

### **Sites web de référence**

Bibliothèque nationale de France : <<http://gallica.bnf.fr>> [février 2007].

Archives littéraires suisses : <<http://www.nb.admin.ch/slb/>> [février 2007].

Bibliothèque nationale du Luxembourg, Centre National de Littérature : <<http://literaturarchiv.lu/fr>> [février 2007].

Bibliothèque royale de Belgique : <<http://www.aml.cfwb.be>> [février 2007].

Bibliothèque nationale du Québec : <<http://www.bnquebec.ca/>> [février 2007].

Le numérique : impact sur le cycle de vie du document, 2004, Montréal.

Archives départementales de Savoie : <<http://www.sabaudia.org>> [février 2007].

Bibliothèque municipale de Lisieux : <<http://bmlisieux.com>> [février 2007].

Centre Flaubert de l'Université de Rouen : <<http://www.univ-rouen.fr/flaubert/>> [février 2007].

Projet IRMA : <<http://www.unites.uqam.ca/irma/>> [février 2007].