

Cahiers **GUT** *enberg*

« LE MONDE SELON ARCANES » ; UN PARADIGME INSTRUMENTAL POUR L'ÉDITION ÉLECTRONIQUE

Éric-Olivier LOCHARD, Dominique TAURISSE

Cahiers GUTenberg, n° 39-40 (2001), p. 89-105.

http://cahiers.gutenberg.eu.org/fitem?id=CG_2001__39-40_89_0

© Association GUTenberg, 2001, tous droits réservés.

L'accès aux articles des *Cahiers GUTenberg*

(<http://cahiers.gutenberg.eu.org/>),

implique l'accord avec les conditions générales

d'utilisation (<http://cahiers.gutenberg.eu.org/legal.html>).

Toute utilisation commerciale ou impression systématique

est constitutive d'une infraction pénale. Toute copie ou impression

de ce fichier doit contenir la présente mention de copyright.

« Le monde selon Arcane » : un paradigme instrumental pour l'édition électronique

Éric-Olivier LOCHARD*
Dominique TAURISSON**

* Université Montpellier 3

** CNRS

Résumé. Cet article décrit un paradigme instrumental destiné à l'édition de connaissances multimédias établies selon des méthodes scientifiques, sous forme de livre papier ou de livre électronique. Après en avoir donné quelques fondements et principes, on présente le processus d'édition comme deux activités de rédaction et publication, appliquées à un monde enregistré dans une base de données. On montre comment l'architecture de ce monde associée à un dispositif de prescriptions rend indépendantes ces deux activités et permet de produire des publications sous différentes formes à partir d'un même travail de rédaction. On présente ensuite la lecture comme l'exploration intentionnelle de ce monde, à l'aide de traitements formulés par le lecteur dans le métalangage de description du monde et exécutés par le livre électronique se comportant alors comme un calculateur.

Abstract. *The World according to Arcane is an operating instrumental paradigm for the electronic edition of scientifically established texts and knowledge, currently being used in several projects of scholarly edition. The world of knowledge is edited in a data base the architecture of which is both generic, so as to be applicable to numerous domains, and simple: any information is a subject of interest, a multimedia document, a relation between subjects or an enrichment. Internal or external documents are enriched with the editing module, independently of the media and the final publication. The reading module offers very powerful procedures to investigate and browse electronic work: typified links inferred by the architecture, sophisticated indexation, dynamic composition of virtual documents, naturally formulated requests, formal treatments, and reading itineraries. The publishing module allows to export information in various formats (HTML, XML, T_EX), to compose paper books, and to produce electronic books in the form of autonomous applications distributed on CD-ROM, DVD-ROM, web site, or data base system.*

« Le monde selon Arcane » est un paradigme instrumental pour l'édition de livres électroniques consacrés à la diffusion de connaissances, éditions savantes, pédagogiques ou documentaires.

Il est l'aboutissement d'un travail de recherche effectué par une équipe que j'ai dirigée, « Modèles informatiques pour l'analyse et l'édition des textes », au sein du Centre d'étude du XVIII^e siècle, CNRS-Université Paul-Valéry¹.

Arcane a été présenté pour la première fois, en 1994, à la Maison des sciences de l'homme à Paris, après avoir été mis à l'épreuve avec l'édition des *Lettres de France* de Denis Fonvizine dans la collection *Archives de l'Est* (dir. G. Dulac, CNRS-Éditions, Fondation Voltaire), la fabrication ayant été effectuée dans l'imprimerie Louis-Jean de Gap, à partir de documents balisés en T_EX.

Il est aujourd'hui utilisé dans plusieurs programmes : l'édition de la *Correspondance complète de Pierre Bayle* (dir. A. McKenna, Fondation Voltaire, Oxford), dont nous venons d'éditer le second volume, une maquette de CD-ROM pour le *Dictionnaire de Port-Royal* (dir. A. McKenna, Champion, Paris), l'inventaire de la *Correspondance de Samuel Formey* (dir. J. Häsel, Champion, Paris), l'édition électronique du *Journal du Chevalier de Corberon* (utilisée pour les illustrations). Des collaborations sont par ailleurs en cours, avec notamment l'institut Constantin-Huygens de la Haye (dir. D. van Vliet) et la bibliothèque universitaire de Leyde (C. Berkvens-Stevelinck).

1. Le livre électronique selon Arcane

À l'origine, « Le monde selon Arcane » a été conçu pour donner une réponse instrumentale au problème de l'adaptation de l'édition aux nouvelles technologies électroniques et informatiques, l'édition étant définie ici comme le processus qui, d'un projet éditorial, aboutit à la diffusion papier ou électronique, en passant par les étapes de rédaction, production et publication. Mais, peu à peu, le livre électronique est devenu une préoccupation majeure, si bien qu'aujourd'hui Arcane est en même temps un modèle fonctionnel du livre électronique et une instrumentation pour son édition.

Dans ce paradigme, un livre électronique est d'abord un livre, objet matériel ou virtuel distribué au lecteur : c'est un ensemble cohérent d'informations établies selon des méthodes scientifiques, destiné à répondre à une attente de lecteurs identifiables et labellisé comme tel par une autorité socio-économique, l'éditeur (tout ensemble d'informations ne constitue pas un livre).

1. Il est intéressant de noter que les difficultés rencontrées pour évaluer ce genre de recherche appliquée et comprendre l'interdisciplinarité au sein d'un laboratoire des Sciences de l'homme et de la société (par exemple une équipe dirigée par un informaticien dans un centre d'étude de la littérature) ont provoqué la dissolution de l'équipe en septembre 2000.

Comme un livre numérique, un livre électronique est donné à lire par l'intermédiaire d'un système électronique (le module de lecture), conçu pour exécuter des procédures dites de lecture, en réponse à des instructions formulées par le lecteur de manière interactive. Si le livre numérique est une « sorte d'assemblage d'un assez grand nombre de pages pré-enregistrées sur un support numérique » comme on peut le lire dans *Le Robert*, le livre électronique est caractérisé par l'architecture de l'information qu'il contient et sa rédaction, essentiellement conditionnées par ces procédures de lecture programmables : affichage de l'information dans des vues, annotations, composition dynamique de documents virtuels, traitements formels (linguistiques ou logiques par exemple), graphiques, statistiques, production de parcours intentionnels de lecture ou encore communication avec d'autres publications électroniques. Si, indépendamment du contenu, on peut décrire la qualité d'un livre électronique par la pertinence et l'efficacité de ces procédures, la question de savoir ce qu'est un « bon » livre électronique est, à notre connaissance, encore ouverte.

Un livre électronique peut prendre la forme d'une application autonome diffusée sur CD-ROM ou DVD-ROM, d'un site Web « passif », ou d'un système de base de données client-serveur (le livre électronique, hébergé dans un SGBDR, est lu à l'aide d'un browser Web ou d'un module de lecture spécialisé).

2. Quelques objectifs

Au cours du travail d'élaboration d'Arcane, nous avons privilégié quelques problématiques :

- inscrire cette instrumentation dans la logique industrielle de l'édition ; nous avons cherché notamment à développer un instrument qui permette de produire, à partir d'un même travail de rédaction, aussi bien des livres papier que des livres électroniques sous différentes formes ; pour la même raison nous avons voulu concevoir un instrument qui favorise le travail collectif et collaboratif ;
- proposer un modèle algorithmique fondé et « naturel » des activités d'écriture et de lecture électroniques, trait d'union entre rédacteur et lecteur, utilisé par l'un et l'autre pour écrire lisiblement et lire de façon intentionnelle et qui puisse résoudre le problème posé par la perte de la linéarité de l'information publiée : rappelons que la non-linéarité est due à la nature multimédia de l'information, à sa structure hypergraphique et à la nécessité d'un calculateur électronique pour la mettre en œuvre ;
- concevoir un instrument ouvert, capable de traiter des sources d'information multimédias externes, d'exporter des informations rédigées sous divers

formats (XML ou T_EX pour la production de livres papier) et de communiquer avec d'autres publications électroniques.

3. Processus d'édition dans Arcane

Le principe fondateur du paradigme est d'articuler la communication entre rédacteurs et lecteurs autour d'un monde de connaissances, le monde de x selon Arcane, décrit par un métalangage opératoire ; ce monde est composé par le rédacteur, puis donné à lire au lecteur pour le découvrir, le comprendre, et le commenter.

L'édition électronique est définie comme un processus coopératif, sans fin, consistant à :

- définir et prescrire l'architecture du monde à éditer, comme instance de la méta-architecture de l'information définie par le paradigme ;
- construire scientifiquement dans une base de données le monde des sujets d'intérêt du projet éditorial, conformément à cette architecture ;
- enrichir ce monde pour le décrire et le structurer, pour l'illustrer et l'interpréter à l'aide de documents multimédias ou de relations et bien sûr pour le représenter ;
- en extraire à des moments choisis des sous-ensembles cohérents pour produire et publier des livres électroniques.

Les quatre temps du processus, prescription, construction, enrichissement et production constituent en pratique la méthodologie Arcane.

3.1. Prescription du monde

Le responsable scientifique de l'édition (plus tard identifié au prescripteur) commence par définir la structure du monde à éditer, l'ensemble des enrichissements que les rédacteurs pourront effectuer et la charte graphique de la représentation des informations ; cette étape correspond en programmation par objet à la construction de classes instanciables comme sous-classe de classes virtuelles définies dans le paradigme ; en pratique, elle permet de configurer l'outil de rédaction en proposant aux différents rédacteurs un cadre de travail commun et cohérent.

3.2. Rédaction de l'ouvrage

Ensuite, les auteurs-rédacteurs construisent et enrichissent le monde conformément aux prescriptions de l'éditeur, sans se préoccuper de la forme du livre qui sera *in fine* produit et distribué au lecteur, en composant dans une base de

données ce que le paradigme appelle un ouvrage électronique : dans l'édition selon Arcane, il est essentiel de faire la distinction entre l'ouvrage électronique rédigé « en sémantique » par les rédacteurs et le livre électronique produit à partir de l'ouvrage, contrairement à l'édition classique où le livre papier est à la fois l'objet rédigé et l'objet distribué.

3.3. Production du livre

Enfin, à des moments déterminés en fonction de la stratégie de diffusion de l'éditeur, le producteur exporte tout ou partie de l'information contenue dans l'ouvrage, en la formatant avec des prescriptions, pour produire un livre électronique sur le support choisi. On remarquera que la production d'un livre ne signifie pas la fin du processus d'édition comme dans l'édition classique : rien n'empêche en principe qu'un même ouvrage électronique puisse donner lieu à plusieurs publications, sur divers supports, pour des publics différents et même à des publications successives ou en continu (Web).

4. Architecture du monde

Le monde se compose essentiellement de sujets, de documents, de relations, d'enrichissements et d'un métalangage pour le décrire.

4.1. Les sujets

Les sujets constituent les centres d'intérêt du projet éditorial, personnes, lieux, ouvrages, événements, thèmes d'étude, notions ou autres, selon les choix de l'éditeur, le principe étant que le monde édité ne parle de rien d'autre que de ses sujets ou à propos de ses sujets : ils déterminent la sémantique effective et objective du monde édité.

Pour le rédacteur, ils sont les candidats de l'indexation sous toutes ses formes et, sur le plan méthodologique, ils facilitent les plans de travail rigoureux et exhaustifs.

Pour le lecteur, les sujets sont des entrées privilégiées pour explorer et lire le monde, la voie d'accès naturelle à une grande partie des informations qu'il contient.

Les sujets sont tous décrits par des champs canoniques qui sont des valeurs typées : un identificateur d'usage (chaîne alphanumérique), une forme normalisée de publication (texte enrichissable), un identificateur universel pour la communication avec d'autres mondes (entier), le créateur du sujet (entier),

un état d'édition (chaîne), un ensemble de formes d'usage équivalentes ou synonymes.

Un sujet appartient généralement à un type, défini par un nom d'usage, un identificateur universel et un ensemble de champs typés (chaîne, texte enrichissable, nombre, date, énumération, sujet et document) qui permettent une description commune de tous les sujets du même type. Les types de sujets et leurs champs sont librement introduits dans le monde par l'éditeur sous la forme de prescriptions.

Les relations permettent d'organiser les sujets dans des structures généralement arborescentes, par exemple pour exprimer l'appartenance.

Les noms des sujets, des types et des champs, de même que les formes équivalentes des sujets font partie du métalangage de description du monde et sont autant d'outils lexicaux mis à la disposition des rédacteurs et des lecteurs pour explorer et interroger le monde.

4.2. Les documents enrichis

Ce sont des informations multimédias destinées à décrire, illustrer, annoter, commenter ou interpréter le monde.

Les documents se répartissent en trois classes : les documents autonomes, les codocuments de document, sujet ou relation et les documents dynamiques. Les codocuments sont une généralisation des « notes de bas de page » : ils enrichissent d'une ou plusieurs informations documentaires multimédias les documents (codocuments de document), mais également les sujets (codocuments de sujet) et les relations (codocuments de relation) ; ils se lisent naturellement « depuis » leur source. Ils ont été pensés pour faciliter la rédaction et la lecture du monde, en introduisant explicitement de l'ordre dans les documents : toute information documentaire dont l'existence dans le monde dépend de celle d'un sujet (biographie d'une personne, exercice d'application d'un sujet-notion) « doit » être un codocument de ce sujet et se lire depuis le sujet ; toute note ou commentaire d'un document est un codocument et se lit depuis ce document.

Les documents dynamiques sont définis par un script à appliquer sur des éléments du monde ; ils sont effectivement calculés et produits par le module de lecture en réponse à une instruction du lecteur, par exemple le document textuel composé de toutes les ancrs d'une indexation par un sujet donné ou un « diaporama » Quicktime obtenu en juxtaposant des documents images enrichis de codocuments sonores.

Sujets : 24 fiches sur 6450

Intit  Corberon1901 DuvLitt 75002

Corberon, Marie Daniel Bourr e, baron de, *Un Diplomate fran ais   la Cour de Catherine II, 1775-1780. Journal intime du chevalier de Corberon, charg  d'affaires de France en Russie*, publi  par L.-H. Labande, Paris, Plon-Nourrit, 1901, 2 vol.

Champs Notice Enregistrer

Champ	Type	Mod	Auteur	Valeur
Forme normale	Texte		DTM	Un Diplomate fran�ais � la Cour de Catherine II,
Compit	Texte		DTM	, du baron de Corberon
Admin	Texte		DTM	
SourceBiblio	Texte			
�uvre	Sujet			
Auteur	Sujet	&	DTM	Corberon M. D. B. de
DateEdition	Date		DTM	1901
LieuEdition	Sujet			
LieuConservation	Enum			
NatureOuvrage	Enum			
EditeurIntellectuel	Sujet			
Langue	Enum		DTM	Fran�ais
Notice	Texte		DTM	Corberon, Marie Daniel Bourr�e, baron de, Un D
Abre�g�	Texte		DTM	Corberon, 1901
FormeLitt	Enum			

CHAMPS

F C D A R 29902 T M 7 24 Cacher

FIGURE 1 – Un sujet dans la vue des champs

Chaque document appartient   un genre (article, illustration, commentaire sonore, etc.) qui lui impose un style de mise en forme et  ventuellement un type avec ses champs.

Arcane distingue op ratoirement le document mat riel multim dia de base, de ses enrichissements cr s explicitement pour le formater, le structurer, le commenter ; un document   lire est alors un document enrichi, d fini comme un document de base interpr t  par un ensemble d'enrichissements. La caract ristique principale de ce mod le, qui est la raison de son choix, est qu'un m me document de base peut donner lieu   plusieurs interpr tations, aussi bien structur les pour obtenir des repr sentations diff rentes (repr sentation de travail, impression sur papier ou visualisation sur browser web par exemple), que s mantiques pour donner   lire ou   r diger des points de vue diff rents.

Une autre de ses qualit s est que ce document de base peut appartenir   d'autres bases documentaires et  tre charg  lorsque le lecteur-r dacteur consulte le document : le document enrichi est alors compos  dynamiquement avec les enrichissements qui eux sont enregistr s dans l'ouvrage.

Un document (de base) est d coup  en sections, essentiellement pour des raisons pratiques, chaque section pouvant  tre vue comme une suite lin aire d'unit s mat rielles, caract res d'un texte, points en couleur d'une image,

points temporels d'un son ou d'un film. Une séquence est définie comme une suite contiguë d'unités.

4.3. Les relations et leurs occurrences

Les relations sont des expressions formelles qui servent à établir des liens sémantiques et dynamiques entre les objets du monde, en particulier entre les sujets ; elles sont librement prescrites par l'éditeur sous la forme d'un nom, d'une liste d'arguments typés et éventuellement d'un identificateur universel.

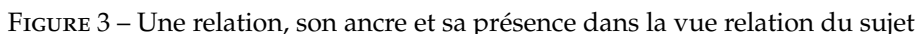
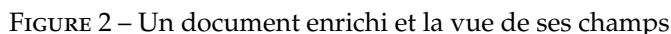
Les occurrences des relations sont généralement créées pour interpréter formellement le contenu d'une séquence d'un document ; on utilise pour cela un enrichissement par relation ancré sur cette séquence. Elles peuvent également être flottantes, pour exprimer une relation objective entre deux sujets, par exemple d'appartenance.

Les relations sont des outils de lecture très efficaces. D'abord parce qu'elles permettent d'augmenter la combinatoire sémantique du monde et donc son exploration ; le lecteur-rédacteur peut par exemple définir opératoirement des termes à l'aide d'une formule d'unification puis utiliser ce terme dans des requêtes (à partir de la relation Lire (lecteur:sujet; ouvrage:sujet; date:date; lieu:sujet), on peut définir une lecture comme un sujet en argument « ouvrage » d'une occurrence de cette relation). Ensuite parce qu'elle permettent de définir des vues de sujet dynamiques (la vue des lectures du sujet x). Enfin parce qu'elles se prêtent à toutes sortes de représentations et traitements (documents dynamiques, statistiques, graphiques, logiques, etc).

4.4. Les enrichissements et leurs occurrences

Ils permettent d'attacher des informations de toutes sortes non seulement aux documents (de base), mais également aux sujets et aux relations : on les utilise pour étiqueter et formater (enrichissements sémantiques), organiser (enrichissements de division ou d'insertion), interpréter et indexer (enrichissements par relation), relier (enrichissements de liens explicites), annoter et codocumenter (enrichissements codocumentaires).

Chaque enrichissement est attaché à une ancre, l'objet enrichi ou une séquence sélectionnée par le lecteur-rédacteur s'il s'agit d'un document ; il peut être associé par le prescripteur à des attributs typés, renseignés au moment de l'enrichissement, à des formats de représentation, des styles typographiques pour la représentation de l'ancre à l'écran ou des balisages attribués utilisés pour l'exportation du document enrichi.



Du point de vue de la lecture, les enrichissements sont généralement activables, par exemple sous la forme d'un lien dynamique d'exploration et de navigation.

Les enrichisseurs, outils de production des enrichissements, sont librement définis par le prescripteur avec leurs éventuels attributs, leurs formats de représentation et les objets sur lequel le lecteur-rédacteur peut les appliquer (types de sujet, genres de document). L'éditeur peut ainsi proposer aux rédacteurs un environnement contextuel de travail cohérent et commode, en particulier par l'usage de chartes graphiques.

4.5. Le métalangage

Le métalangage du monde se compose du métamétalangage Arcane, du métalangage défini par le prescripteur (types de sujets, champs de type et énumérations, relation et arguments, genres des documents, enrichisseurs avec leurs attributs), des éléments lexicaux introduits dans le monde par les rédacteurs (identificateurs d'usage notamment).

Le but premier de ce métalangage est de permettre au lecteur-rédacteur d'identifier « en langue » le contenu du monde édité et son architecture et donc de le parcourir de manière active et intentionnelle.

La seconde raison d'être du métalangage est d'ordre pratique : le module de lecture étant un système programmable, tout ce qui favorise l'énonciation de requêtes facilite en principe l'utilisation du système, l'idéal étant de pouvoir piloter ce module à la voix ; encore faut-il pour cela intégrer au métalangage des procédures de lectures pertinentes et opératoires, qui restent à concevoir, expérimenter et valider.

4.6. Liaisons induites par cette architecture

Cette architecture établit implicitement des liens généralement bidirectionnels entre les différents objets du monde, liens de description (par exemple un champ de type sujet est un lien virtuel entre deux sujets, de même une relation ayant des sujets pour arguments) ou liens par enrichissement (une indexation est un lien virtuel entre un document et un sujet, une codocumentation de sujet est un lien entre un sujet et un document). Ces liens sont parfaitement identifiables par le métalangage et sont donc autant de chemins d'exploration que le module de lecture doit facilement permettre au lecteur d'emprunter.

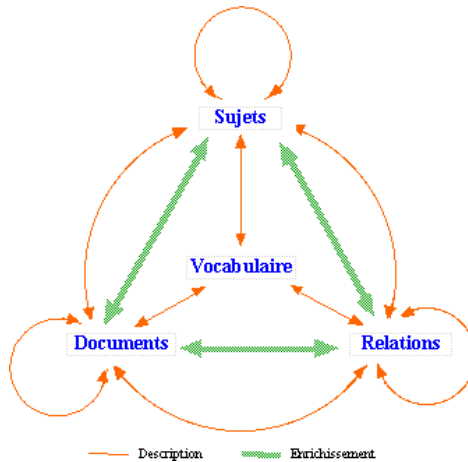


FIGURE 4 – Graphe de l'architecture Arcane

5. Représentations et lectures du monde

Le paradigme Arcane ne décrit pas seulement la méta-architecture des informations du monde. Il donne également un ensemble de principes et/ou de procédures pour le représenter et le lire.

5.1. Vues

Dans Arcane, les informations sont toujours présentées à la lecture dans une vue. Les listes d'objets sont présentées dans des vues-en-liste activables : le lecteur peut facilement provoquer l'affichage des objets présentables de cette liste.

Un sujet est présentable dans quatre vues (au moins) : la vue de ses champs (figure 1), la vue de ses codocuments, la vue des occurrences de relations dont il est l'un des arguments (figure 3) et la vue des ancrs des enrichissements concernant le sujet (indexation notamment).

Un document enrichi est présentable dans trois vues : la vue de ses champs et la vue d'une section de son contenu dans une interprétation donnée (la section du document de base est présentée avec la représentation prescrite des ancrs composant l'interprétation), enfin la vue des sujets associés à ce document par une interprétation.

L'occurrence d'une relation est présentable dans deux vues : la vue de ses arguments et la vue de ses codocuments.

5.2. Liens dynamiques

Nous avons déjà mentionné les liens dynamiques induits par la méta-architecture Arcane : liens de description et liens d'enrichissement. Le paradigme Arcane impose évidemment que ces liens soient facilement activables par le lecteur, éventuellement avec le commentaire attribué.

5.3. Requêtes

Les requêtes sont des énoncés définissant sans ambiguïté un ensemble d'objets du monde (les sujets dont le champ..., les documents indexés par un sujet dont...); par principe, ces requêtes doivent être formulables dans le métalangage et exécutables par le module de lecture pour produire et présenter la liste de tous les objets du monde répondant à la définition.

5.4. Traitements

Les traitements de lecture sont des procédures exécutées par le système en réponse à des demandes formulées d'une manière ou d'une autre par le lecteur. Nous avons déjà évoqué les documents dynamiques définis par des scripts à appliquer à des éléments du monde. Le travail de caractérisation de ces scripts reste encore à faire, de même que le travail sur des traitements graphiques, statistiques, linguistiques ou logiques. Comme Arcane a été conçu pour pouvoir aisément exporter-publier des informations du monde dans divers formats, on imagine facilement que ces traitements peuvent être faits en dehors d'Arcane, à l'aide d'outils spécialisés. Le problème se restreint donc à la description des traitements élémentaires standard qu'un système Arcane devrait au minimum proposer au lecteur.

5.5. Parcours de lecture

Parmi les procédures et principes que nous avons déjà présentés, certains peuvent être naturellement considérés comme des parcours de lecture : les vues des sujets donnent à lire ce que le monde dit d'essentiel sur ces sujets, les requêtes permettent de composer des listes à parcourir, etc.

Arcane propose un dispositif supplémentaire : le parcours intentionnel de lecture ; il consiste à écrire un scénario « linéaire » composé d'une suite commen-

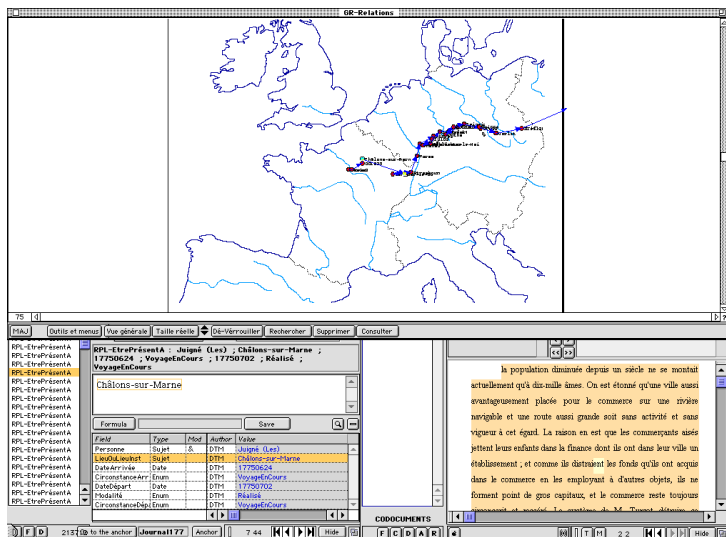


FIGURE 5 – Représentation graphique de relations

tée d'objets du monde dans des vues, de le composer dans l'espace (choisir la fenêtre pour afficher chacune des vues) et dans le temps (synchroniser le moment et le temps d'affichage) et le proposer au lecteur pour une lecture passive (le scénario se déroule automatiquement) ou active (le lecteur déclenche lui-même la vue suivante, à partir de laquelle il peut toujours naviguer). Ces parcours intentionnels peuvent aussi bien être des initiations ou présentations du monde édité, que des exercices demandés à des lecteurs étudiants autour d'un thème donné.

6. Publication

Dans Arcane, la publication est une initiative de l'éditeur ; elle consiste à choisir les éléments enrichis du monde qui vont composer le contenu de la publication, extraire ces éléments selon des protocoles spécifiques et produire un livre papier ou électronique.

Dans le cas d'une publication du genre dictionnaire, documentation ou encyclopédie, les objets publiés seront des sujets, les documents servant essentiellement à codocumenter ceux-ci (mis à part les éléments des parcours intentionnels de lecture) ; dans le cas d'une édition critique, ce seront au contraire des documents, les sujets étant surtout utilisés pour les index ; dans le cas d'une

édition pédagogique les choix sont *a priori* beaucoup plus ouverts : notre expérience d'éditions de ce type est encore trop limitée pour proposer un modèle raisonné de livre électronique.

En pratique une publication revient à sélectionner des éléments de la base de données, à décrire leur agencement par un script, prescrire les enrichissements à prendre en compte et enfin appliquer un traitement à chacun de ces éléments, essentiellement le codage des informations de base et le balisage des enrichissements, pour produire un ensemble de fichiers dans un format *ad-hoc* : fichiers PostScript ou T_EX pour un livre papier, fichiers HTML pour un site Web, fichiers XML pour une application autonome ou une base de données avec module de lecture propre à l'éditeur. Le cas d'une application autonome ou base de données avec un module de lecture Arcane est un peu particulier : si on supprime les informations à ne pas publier, l'ouvrage électronique est un livre électronique.

6.1. Traitement d'un sujet ou d'une relation

La production d'un fichier à partir de sujets ou de relations ne pose pas de problème particulier, si on s'en tient uniquement aux champs et aux arguments (et si on reste dans le contexte de ces journées) : on a affaire à des structures arborescentes qu'on peut produire avec une grammaire algébrique ; pour décrire cette grammaire par une DTD (*Document Type Declaration*) SGML, il suffit pratiquement de faire les choix entre balise et attribut (nous avons choisi le principe de réduire les attributs au minimum) et de prescrire une balise pour chaque champ et chaque argument. Le problème auquel nous venons de faire allusion concerne l'identification des objets dans la communication avec d'autres mondes (insertion d'un livre électronique dans une collection électronique par exemple). Le principe retenu par Arcane est d'utiliser un identificateur universel pour tous les sujets, les relations et les documents de base universels. Ce principe, pour élégant qu'il soit (nous le mettons en œuvre dans la communication des mondes que nous éditons), n'a pas encore fait la preuve de sa viabilité, si on en juge par les politiques poursuivies en la matière par les grandes institutions savantes.

6.2. Traitement d'un document

Le principe retenu par Arcane consiste à coder le document de base « de gauche à droite », en balisant au fur et à mesure les ancres des enrichissements à prendre en compte. Rappelons que ce balisage est défini dans les prescriptions, en fonction du format de production (balises T_EX, XML, HTML, etc.), de même que le traitement des attributs associés aux enrichissements ; ceux-ci peuvent être traduits en attributs SGML ou en arguments de commandes T_EX.

Ce faisant, il faut faire un certain nombre de vérifications en fonction du format choisi. Pour XML et HTML il faut vérifier l'algébricité du balisage, puisqu'aucune contrainte de ce type n'a été imposée au rédacteur. La pratique montre que, comme l'ont probablement supposé les concepteurs de SGML, les enrichissements publiables forment la plupart du temps des expressions bien parenthésées. Par ailleurs il faut transformer les prescriptions de formatage des enrichissements en feuilles de style CSS (*Cascading style sheet*). Pour T_EX, il faut contextualiser le codage (par exemple celui du caractère retour-chariot). Certaines contextualisations sont automatiquement effectuées par Arcane (codage d'un codocument dans une note de bas de page par exemple); le producteur peut cependant définir explicitement des contextes de codage par des prescriptions d'édition (codage de vers, de citations par exemple). Bien entendu, Arcane est conçu pour favoriser l'usage de styles propres à chaque ouvrage.

6.3. Publication et politique éditoriale

En principe, la production d'un livre électronique n'interrompt pas l'activité d'édition; elle peut se poursuivre pour d'autres productions, avec la même équipe de rédacteurs ou avec une autre. De plus, l'architecture modulaire de l'instrument Arcane, composé de différents modules logiciels appliqués à une même base de données (prescription, rédaction, production, lecture, administration), permet de bien distinguer les différentes phases de travail, de les organiser efficacement et de mettre en œuvre des politiques éditoriales très élaborées. Encore faudrait-il que les éditeurs institutionnels, prenant des distances avec une frénétique stratégie de présence sur le Web à tout prix, deviennent pour le livre électronique ce qu'ils ont longtemps été pour le livre papier, des éditeurs.

7. Communication entre éditions électroniques

Arcane a été conçu dès l'origine comme un instrument ouvert, pour permettre la communication avec d'autres mondes édités ou d'autres éditions : nous avons déjà vu comment le rédacteur pouvait travailler sur des documents de base externes, appartenant à d'autres éditions électroniques; et nous venons de voir comment publier l'information du monde dans un format XML (ou une de ses variantes) supposé devenir un protocole logique standard de communication.

Une autre forme de communication intéressante qui est envisagée est d'utiliser le module de lecture pour interroger dynamiquement d'autres éditions

électroniques capables de communiquer avec Arcane. Cette interrogation suppose d'une part, que le module soit capable d'énoncer la requête dans un langage compréhensible, type SQL et d'autre part, de disposer d'un algorithme d'identification des composantes de ces requêtes, en particulier les types de sujets avec tout ou partie des champs associés, les sujets eux-mêmes, les documents de base et les relations. À défaut d'expérimentation en la matière, Arcane a prévu comme solution générale à ces questions, le métalangage à partir duquel on pourra développer un traducteur et un identificateur universels pour toutes les informations communicables.

Quand à la possibilité d'importer « automatiquement » dans un monde Arcane l'information provenant d'autres éditions, la capacité d'Arcane à lire du code XML ne suffit pas : il faut évidemment que l'architecture de l'information importée soit compatible avec l'architecture Arcane, autrement dit que sa DTD (ou une de ses restrictions) soit une DTD Arcane.

Bibliographie

- [1] BALPE Jean-Pierre, LELU Alain, IMAD Saleh, *Hypertextes et hypermédias*, Paris, Hermès, 1995.
- [2] BOLTER Jay David, *Writing space : the computer, hypertext, and the history of writing*, Lawrence Erlbaum Associates, 1991.
- [3] CHARTIER Roger, *Forms and meanings : Texts, performances and audiences from Codex to Computer*, Philadelphia, University of Pennsylvania Press, 1995.
- [4] DACOS Marin, « Prestige du papier et avenir du numérique. Recherches ruralistes et édition électronique en 1999 », *Ruralia*, Revue de l'Association des ruralistes français, n° 4, 1999.
- [5] DARNTON Robert, *An Early Information Society*, <http://www.indiana.edu/~ahr/darnton/about.html>
- [6] DELANY Paul, Landow, George P. (dir.), *Hypermedia and Literary Studies*, Cambridge, MIT, 1991.
- [7] FERRAND Nathalie (dir.), *Banque de données et hypertextes pour l'étude du roman*, Paris, PUF, 1997.
- [8] FINNERAN Richard (dir.), *The Literary Text in the digital age*, Ann Arbor, University of Michigan Press, 1996.
- [9] GAGGI Silvio, *From Text to Hypertext : decentring the subject in fiction film, the visual arts, and electronic media*, Philadelphia, University of Pennsylvania Press, 1997.

-
- [10] GOODY Jack, *La Raison graphique : la domestication de la pensée sauvage*, Paris, Minuit, 1979.
 - [11] JOHNSON Steven, *Interface culture : how new technology transforms the way we create and communicate*, New York, Harper Collins, 1997.
 - [12] LANDOW George P., *Hypertext. The Convergence of contemporary critical Theory and technology*, Johns Hopkins University Press, 1992.
 - [13] — *The Digital Word : Text-Based Computing in the Humanities*, MIT, 1993.
 - [14] — *Hyper/Text/Theory*, Baltimore, Johns Hopkins University Press, 1992.
 - [15] — *Hypertext 2.0*, Baltimore, Johns Hopkins University Press, 1997.
 - [16] LÉVY Pierre, *Les technologies de l'intelligence. L'avenir de la pensée à l'ère informatique*, Paris, La Découverte, 1990, rééd., Paris, Le Seuil, 1993.
 - [17] — *Cyberculture* (Rapport au Conseil de l'Europe), Paris, Odile Jacob, 1997.
 - [18] — *L'intelligence collective. Pour une anthropologie du cyberspace*, Paris, Maspero, 1981, rééd., Paris, La Découverte, 1997.
 - [19] LOCHARD Éric-Olivier, « TOPOSATOR et ARCANE : la nouvelle technologie dans le contexte des thésaurus et des éditions critiques », *Literary and linguistic computing*, Oxford University Press, vol. 10, n° 2, 1995.
 - [20] LOCHARD Éric-Olivier « À propos de TOPOSATOR », colloque SATOR de Winnipeg, *Utopie et fictions narratives*, éd. Michel Bareau et Santé Viselli, Parabasis, Altapress, 1995, p. 257-268.
 - [21] LOCHARD Éric-Olivier et TAURISSON Dominique, « Correspondances, réseaux, édition électronique », *La Plume et la Toile. Pouvoirs et réseaux de correspondance dans l'Europe des Lumières*, textes réunis par Pierre-Yves BEAU-REPAIRE, Arras, Presses de l'université d'Artois (à paraître 2001).
 - [22] McLuhan Marshall, *La Galaxie Gutenberg. La genèse de l'homme typographique*, Montréal, HMH, 1967.
 - [23] SNYDER Ilana, *Hypertext : the electronic labyrinth*, New York, New York University Press, 1997.
 - [24] *Surfaces, electronic Journal*, Presses de l'Université de Montréal, dir. Jean-Claude GUÉDON ; « Humanités et informatique : qui a les commandes ? », vol. 8, 1999.