Structuration des manuscrits : Du corpus à la région

Aurèle Crasson, Jean-Daniel Fekete

Structuration des manuscrits : Du corpus à la région

Aurèle Crasson[1] – Jean-Daniel Fekete[2]

[1] ITEM
CNRS UMR 8132
45, rue d'Ulm
F-75005 Paris
Aurele.Crasson@ens.fr

[2] INRIA Futurs/LRI
Bât 490
Université Paris-Sud
F-91405 Orsay Cedex
Jean-Daniel.Fekete@inria.fr

Résumé *: La gestion de manuscrits en tant que document numérique n'a jamais été convenablement traitée : images pour certains, documents textuels pour d'autres. Nous proposons une représentation triple : image/scripto-graphique/textuelle qui permet d'exprimer pleinement les spécificités de l'objet manuscrit. Nous montrons comment cette représentation peut être exploitée pour l'analyse et la navigation dans des corpus de manuscrits littéraires modernes.*

Mots-clés *:* manuscrit, document numérique, représentations multiples, transcription, représentation multi-échelle.

1 Introduction

Le manuscrit est un objet complexe : à la fois textuel, graphique et topographique. Il a toujours échappé aux tentatives de descriptions numériques car il se plie mal à une structuration rigoureuse. Certains manuscrits, relativement réguliers et propres, peuvent être décrits numériquement de manière similaire à un document textuel, avec les spécificités décrites par [LEC'98] [GUS'99] [FEK'99]. Une des caractéristiques des manuscrits d'auteur est d'être un document non stabilisé qui lie de manière entrelacée et inséparable des composantes textuelles, graphiques et spatiales. A ces composantes s'ajoute une dimension chronologique, qui renvoie au principe de l'écriture en processus. L'absence de linéarité du texte « entrain de s'écrire » (mais cependant figé dans une certaine forme sur son support) renvoie quant à elle à une représentation très différente d'un document textuel stabilisé ou imprimé.

Dans cet article, nous décrivons un travail en cours sur la description, l'analyse et la représentation de documents numériques issus de manuscrits littéraires (Flaubert, Proust, Valéry, Jabès).

1.1 Dualité texte/image

Quelque soit l'approche littéraire, linguistique ou sémiotique de ce document, parce ce qu'il s'agit d'un manuscrit d'auteur, le premier regard du généticien du texte se porte sur la trace écrite et le récit. Il n'en demeure pas moins que les informations graphiques, spatiales et codicologiques jouent un rôle indispensable pour appuyer des hypothèses de chronologie ou pour retrouver des séquences textuelles.

Le manuscrit serait donc un support d'écriture qui a fixé un ensemble de traces scripto-graphiques constituées en strates temporelles. Chaque feuillet (ou autre unité distincte) numérisé renvoie à une image dans laquelle se placent les informations textuelles, graphiques et spatiales qu'il est possible de décrire individuellement et grâce à des liens typés. Chacune de ces unités, à l'issue de l'étude génétique est sensé être clairement localisée dans l'ensemble que forme, ce que les généticiens appellent plus explicitement, un avant-texte (tout élément classé d'un manuscrit qui précède le texte définitif).

Une page de manuscrit accentue l'ambiguïté entre le texte représenté et l'image globale du document qui inclut d'autres informations. Le fait que, numérisée elle ne puisse pas être enregistrée autrement que sous forme d'image (il n'y a pas encore de programme qui reconnaisse l'écriture manuscrite), oblige à s'intéresser à d'autres types d'informations que le texte. Par ailleurs, le fait qu'il s'agisse d'écriture manuscrite, parfois difficile à déchiffrer oblige le généticien à transcrire, et à multiplier ainsi les représentations du document source.

Cela pose alors un problème de structuration.

L'analyse de corpus manuscrit est conditionnée par cette étape de transcription qui, outre le fait qu'elle permet de lire plus facilement le texte, donne aussi une représentation d'une page manuscrite qui la plupart du temps est interdite à la reproduction.

Les modalités de transcription n'ont jamais été normalisées. S'il est entendu que la transcription se doit d'être la plus objective possible, les corpus ont en réalité une énorme influence sur les formes de rendu. Trois types de transcriptions que l'on peut classer de la plus objective à la plus interprétative permettent jusqu'à présent de traduire le manuscrit :

- la transcription diplomatique : elle « photographie » le document en rapportant, avec les outils qui le permettent, malgré leurs limites, tous les événements du manuscrit ;
- la transcription linéarisée : elle ne prend en compte ni les données topographiques ni les données « vectorielles » ; (elle « aligne » toutes les données textuelles en insérant les corrections) ;

– la transcription chronologisée qui empile selon un axe temporel, les périodes d'écriture. Ce dernier type, développé par Jean-Louis Lebrave, linguiste, est très peu utilisé compte tenu de la part importante d'interprétation qu'il suppose.

Avec les technologies numériques on pourrait penser qu'une transcription du manuscrit n'a pas d'utilité puisqu'on peut se référer à l'image numérique ; c'est vrai et faux. Cela répond surtout à un objectif de recherche, qu'il soit provisoire – relever des types de graphiques répétitifs et vérifier si l'hypothèse est vraie sur l'image – ou destiné à servir dans une publication, la citation d'une partie du texte pour illustrer une interprétation, par exemple. Dans le premier cas où le contexte global de l'image et de la transcription est nécessaire parce qu'il s'agit de vérifier des indices potentiels, la représentation photographique de la transcription est nécessaire ; dans l'autre où il ne s'agit en définitive que de se référer à du texte, la représentation linéaire est suffisante.

D'un autre point de vue, plus matériel, l'utilisation d'un traitement de texte même enrichi de description-types ne permet pas l'affichage de toutes les particularités de l'image manuscrite ; notamment les orientations de textes, certains graphismes etc. Il en résulte qu'en transcrivant même le plus précisément possible, on perd ou on ajoute des informations. De là l'intérêt de pouvoir coupler sur un même plan visuel la représentation image et la transcription.

2 La gestion numérique de manuscrits

Plusieurs autres projets ont proposés des méthodes et codages pour décrire les manuscrits. La « Text Encoding Initiative » (TEI) [SPE'02] décrit plusieurs méthodes pour transcrire des sources qui peuvent être manuscrites. Nous avons nous-même utilisé TEI pour transcrire des manuscrits historiques du 16e siècle [FEK'99]. TEI fonctionne bien pour les manuscrits relativement propres où le texte est stable mais ne convient plus lorsque les phénomènes paratextuels prolifèrent, comme c'est le cas dans les manuscrits littéraires modernes ou dans des brouillons. L'expérience montre que l'application de ces descriptions aux manuscrits est bien plus complexe.

Même pour des documents manuscrits propres, certains travaux ont dû développer un environnement spécialisé d'édition critique, comme le projet Bambi [BOZ'97]. Cependant, la façon dont les manuscrits sont décrits dans Bambi se rapproche énormément de la TEI. L'environnement permet de manipuler le document structuré et enrichi de manière plus conviviale, mais sans amélioration structurelle notable. La fonctionnalité la plus importante apportée par Bambi est la mise en relation automatique du texte et de son image. Une fois le texte transcrit manuellement - la reconnaissance automatique de l'écriture manuscrite étant illusoire -, Bambi peut calculer un lien entre la transcription et son image. Ce calcul suppose que le manuscrit soit propre et régulier. Il cherche pour chaque mot de la transcription un bloc d'image ayant les caractéristiques similaires au mot écrit. La transcription produite par Bambi est donc liée à son image, ce qui nous semble fondamental.

Lecolinet et Robert ont aussi travaillés sur un système de mise en relation du texte et de son image en codant le résultat au format TEI. Leur système ne distingue encore que deux représentations pour le manuscrit : son image et sa structure textuelle. Le niveau scripto-graphique n'est pas explicite et pose des problèmes non résolus pour les manuscrits modernes.

Au-delà de la représentation textuelle, le projet d'édition des multiples versions du Chevalier de la Charrette engagé par Princeton en 1997 a abordé le problème des graphies spécifiques dans les manuscrits par la définition d'entités SGML spécifiques à chaque graphie. Il s'agit là d'une codification des pratiques manuscrites. La transcription qui en résulte est très lourde et a nécessité une étude initiale de tous les signes, avec des décisions arbitraires sur le regroupement de signes sous une même classe nommée par une entité. C'est pourtant le début d'un codage scripto-graphique puisque les multiples graphies sont un peu exprimées. Le reste du projet suit le codage TEI. Cette approche est aussi utilisée par l'école des Chartes à Paris.

Dans une autre direction, une extension importante à TEI a été élaborée par le projet européen Master pour décrire les manuscrits. Cependant, Master ne décrit pas la transcription d'un manuscrit mais ses caractéristiques de catalogages. Rien de particulier n'a été ajouté pour les transcriptions. (Peter Robinson Canterbury Tales). Jusqu'à récemment, le consortium TEI n'a pas abordé le problème spécifique du codage des manuscrits, mais il s'y est engagé récemment.

3 Structure d'un manuscrit numérique

3.1 Modèle en couches

La numérisation n'est pas un but en soi, elle repose sur l'idée qu'un document sur support numérique sera plus facile à consulter et à analyser, cette analyse étant destinée à publication par exemple. Les premiers projets de numérisation de manuscrits tels Bambi étaient orientés vers une finalité claire : l'analyse historique pour une publication. En réalité, le développement des recommandations de la TEI a montré que, lorsqu'un document est numérisé, plusieurs personnes ou projets veulent le réutiliser à des fins parfois très différentes. Dans notre propre expérience de numérisation des lettres de rémission de la Renaissance, nous avons été sollicité par plusieurs chercheurs qui désiraient analyser ces lettres de rémission à des fins très variées comme l'analyse de la langue française du 16e siècle. Nous concevons la numérisation comme une étape dans un processus d'analyse créant du sens et des structures à partir de sources. Ce processus organise une hiérarchie d'annotations ou d'analyse visant un ou plusieurs buts, certaines étapes étant génériques tandis que d'autres sont plus spécifiques. Par exemple, pour être lu, la plupart des manuscrits doivent être transcrits. La transcription constitue un premier niveau d'annotation qui est générique. A ce niveau, il est aussi possible de faire des

descriptions de manuscrits ; c'est ce que décrit l'extension de TEI « Master » [ESP'99]. Ces premiers niveaux peuvent devenir des ressources pour des analyses de plus haut niveau, qu'elles soient linguistiques, historiques, sociologiques, etc. Certaines de ces analyses peuvent reposer sur des traitements automatiques, comme la lemmatisation pour l'analyse linguistique. D'autres traitements peuvent produire des indexs ou faciliter la création de glossaires etc. Tous ces éléments peuvent à leur tour faire l'objet de réutilisation pour des articles savants, des éditions en ligne ou des mises en valeur muséographiques.

Nous concevons donc l'édition électronique très généralement comme un processus de création de niveaux d'annotations et d'analyse de plus en plus abstraites.

Du point de vue du modèle en couches, le manuscrit en présente trois différentes : l'image, la décomposition scripto-graphique de l'image et la structuration textuelle. Notre principale contribution dans cet article est de mettre en évidence cette représentation pivot des manuscrits qu'est le niveau scripto-graphique. Si le niveau textuel a été bien étudié et décrit par les recommandations de la TEI, le niveau scripto-graphique reste encore inexploré. D'un point de vue abstrait, c'est pourtant, en partie, à travers la relation qui lie ces informations à la topographie, que se constitue le manuscrit.

3.2 Découpage en régions

La description scripto-graphique d'un manuscrit est une transcription structurée qui repose sur l'unité du feuillet manuscrit. Ce feuillet est considéré comme un ensemble de régions : zones graphiques connexes contenant récursivement des éléments scripto-graphiques (Fig. 1). Ces régions peuvent parfois se superposer ; elles contiennent des unités de lectures reconnues par le transcripteur. Une étude informelle menée sur une dizaine de personnes nous a montré que cette décomposition était extrêmement stable entre spécialistes. Chaque région est définie par une enveloppe polygonale simple et une orientation principale, pas toujours horizontale ou verticale.

Chaque région est composée de régions de dimension variable, d'éléments textuels et d'éléments graphiques. La description textuelle est calquée sur la TEI tandis que les éléments graphiques se conforment à SVG [FER'03]. Les primitives graphiques SVG ont le statut d'un caractère ou d'un élément intra-textuel ou para-textuel. Par exemple ; certains auteurs utilisent des abréviations de manière systématique. Ces abréviations seront alors codées textuellement à l'aide de la balise <abbr> de TEI. D'autres abréviations peuvent être moins fréquentes et ambiguës. Elles doivent alors être codée comme une zone graphique ayant un statut de mot ou abréviation mais dont le dessin est sur une région particulière de l'image et éventuellement reproduite graphiquement en SVG dans la transcription. Enfin, certains signes du manuscrit, dont le sens n'est pas totalement clair, doivent être liés à leur image et éventuellement transcrits graphiquement. Cette reproduction n'est pas indispensable tant que la liaison transcription/image existe mais, l'ajout du graphique rend conforme la transcription diplomatique et permet, au cours de la lecture, de mémoriser les « évènements » non langagiers.

Des éléments graphiques para-textuels sont nombreux dans les manuscrits d'auteur : une ligne verticale ou horizontale, un « becquet » indiquant l'insertion d'une note marginale entre deux mots d'une autre zone de texte, des dessins, des accolades etc. Ces signes plus ou moins faciles à interpréter, sont parfois sujets à débat. Ils doivent donc être représentés en tant que structure bien que leur fonction ne soit pas toujours évidente. Encore une fois, une simple liaison à l'image pourrait suffire structurellement pour indiquer l'existence d'un phénomène graphique intéressant, mais sa représentation graphique simplifiée en SVG permet de générer une présentation visuelle de la structuration scripto-graphique (une transcription diplomatique) et d'analyser cette transcription hors contexte, ce qui est important.

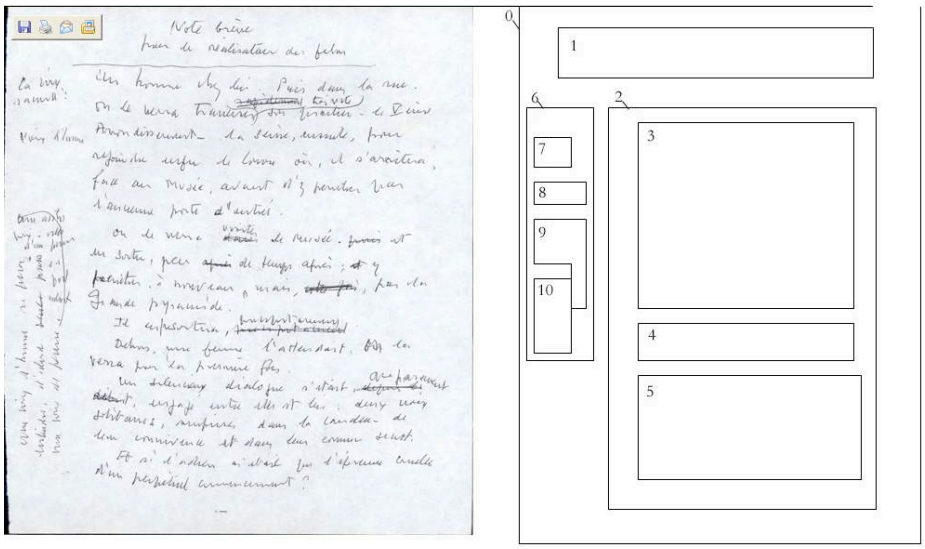

Fig. 1 : Découpage d'un feuillet manuscrit en régions

```
<folio id="folio-18v"
 rend="image(jabs/images/manuscrits/m18versog.jpg) width(1008) height(1134)">
 <fragment id="folio-18v-1"  rend="rect(272 17 741 119)">
  <l id="folio-18v-1-1">Note bréve</l>
  <l id="folio-18v-1-2">pour le réalisateur du film</l></fragment>
 <fragment id="folio-18v-2-30"  rend="rect(15 122 146 228)">
  <l id="folio-18v-2-31">La voix</l>
  <l id="folio-18v-2-32">raconte :</l></fragment>
…
 <fragment id="folio-18v-2"  rend="rect(180 130 936 461)">
  <l id="folio-18v-2-1">Un homme chez lui. Puis dans la rue.</l>
  <l id="folio-18v-2-2">on le verra traverser
   <add place="supralinear">
    <del id="folio-18v-2-3">rapidement</del>
    <add id="folio-18v-2-4">trés vite</add>
   </add> son quartier - le Véme</l>
  <l id="folio-18v-2-5">Arrondissement - la seine, ensuite pour</l>
  <l id="folio-18v-2-6">rejoindre enfin le louvre oé, il s'arrétera,</l>
  <l id="folio-18v-2-7">face au Musée, avant d'y penetrer par</l>
  <l id="folio-18v-2-8">l'ancienne porte d'entrée.</l></fragment>
…
 <fragment id="folio-18v-2-33"  rend="rect(179 473 916 664)">
  <l id="folio-18v-2-9">on le verra
   <del id="folio-18v-2-10">dans</del>
   <add id="folio-18v-2-11" place="supralinear">visiter</add>
   le Musée <del id="folio-18v-2-12">puis</del> et</l>
  <l id="folio-18v-2-13">en sortir, peu <del>aprés</del> de temps aprés ; <del>et</del> y</l>
  <l id="folio-18v-2-14">penétrer, à nouveau, mais, <del id="folio-18v-2-15">cette fois</del>,
   par la</l>
  <l id="folio-18v-2-16">Grand pyramide.</l></fragment>
 <fragment id="folio-18v-2-34"  rend="rect(169 659 974 1053)">
  <l id="folio-18v-2-17">Il en ressortira,
   <del id="folio-18v-2-18">précipitement</del>
   <add id="folio-18v-2-19" place="supralinear">précipitament</add></l>
  <l id="folio-18v-2-191">Dehors, une femme l'attendait. On la</l>
  <l id="folio-18v-2-20">verra pour la premiére fois.</l>
  <l id="folio-18v-2-21">un silencieux dialogue s'était,
   <del id="folio-18v-2-22">depuis le</del>
   <add id="folio-18v-2-23" place="supralinear">auparavant</add></l>
  <l id="folio-18v-2-24">
   <del id="folio-18v-2-25">début</del>,
   engagé entre elle et lui : deux voix</l>
  <l id="folio-18v-2-26">solitaires, surprises dans la candeur de</l>
  <l id="folio-18v-2-27">leur connivence et dans leur commun secret.</l>
  <l id="folio-18v-2-28">Et si l'adieu n'était que l'épreuve cruelle</l>
  <l id="folio-18v-2-29">d'un perpétuel commencement ?</l></fragment>
</folio>
```

Fig. 2 : Codage XML scripto-graphique du feuillet 18 verso du manuscrit de Jabès « Cela a eu lieu »

Comme dans les recommandations de TEI, chaque élément de notre description scripto-graphique peut être enrichi sémantiquement ; il est donc possible d'exprimer directement la fonction de chaque signe au niveau scripto-graphique. Cependant, nous pensons que la résolution des fonctions appartient plutôt à la couche supérieure d'analyse. Concrètement, cela signifie que si une description scripto-graphique est trop annotée, sa réutilisation dans un autre contexte nécessite le retrait des annotations spécifiques. C'est ce qui se produit aujourd'hui avec les documents textuels codés avec TEI.

3.3 Liaison XML / Image

Pour lier la description scripto-graphique à son image, nous utilisons des références externes (Fig. 2). Le monde XML dispose des XPointers/XLink pour exprimer les relations entre documents XML mais rien n'est défini pour exprimer les relations XML/image. Nous avons donc précisé une syntaxe spécifique pour palier ce manque important. Cette syntaxe utilise les spécifications géométriques de SVG pour exprimer des régions d'intérêt dans une image.

Il faut rappeler que le travail réalisé pour spécifier XPointers/XLink était inspiré des pointeurs étendus définis par TEI en SGML (P3). Ces pointeurs étendus définissaient aussi les liaisons avec des documents externes, image ou vidéo. Cette possibilité n'a pas été reprise dans le monde XML et c'est une des inspirations initiales de notre travail [FEK'98].

3.4 Structuration des régions

En plus des régions, éléments textuels et éléments graphiques, un feuillet a une structure. Celle-ci doit être déterminée sur un corpus car les relations entre régions dépassent les frontières de feuillets (chevauchement de textes sur deux feuillets par ex.). La génétique textuelle s'appuie sur deux axes de temporalités : le syntagmatique (progression du récit) et le paradigmatique (réécritures de mêmes séquences). La description scripto-graphique constitue les bases de structuration d'un manuscrit tandis qu'un document de

plus haut niveau décrit la structuration sur les deux axes du corpus. Ce document est conforme aux recommandations de TEI mais fait référence à des éléments appartenant à des descriptions scripto-graphiques. C'est à ce niveau que se décrivent les phénomènes génétiques typologisables. Par ailleurs, des fonctions, telles que «lien d'insertion» ou «groupement», sont données aux éléments graphiques pertinents.

4 Représentations, analyses et outils

La description sur trois niveaux d'un manuscrit n'est pas juste un exercice de style ou un besoin de structuration. Nous avons conçu et réalisé des outils pour exploiter des manuscrits codés à des fins d'analyse, de présentation ou d'exploration. Il s'agit d'outils d'affichage de feuillet avec mise en évidence des régions, de recherche de concordances entre feuillets transcrits et de représentation et navigation dans un dossier génétique.

4.1 Affichage de feuillet en liaison avec les régions

Une feuille de style XSLT, liée à l'image, permet de rendre la forme scripto-graphique représentable en HTML. Un outil d'affichage permettant de montrer une image avec un fisheye centré sur une région d'intérêt (Fig. 3) a été conçu en relation avec cela. Ces représentations améliorent notablement la lecture du document. Si la superposition texte/image est un dispositif de contrôle utile, la version HTML, transcription diplomatique très précise, est un mode standard de consultation.

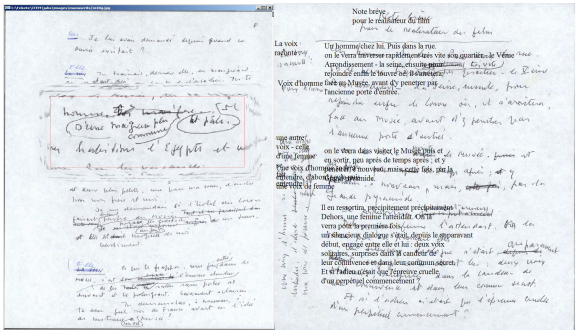

Fig. 3 : Visualisation fisheye d'un manuscrit et visualisation superposée image/texte placé, montrant le lien automatique texte/image du texte

4.2 Recherche de concordances inter-feuillets

Une fois les transcriptions de feuillets réalisées, le travail génétique consiste à établir des relations entre diverses unités et à classer les feuillets sur les axes paradigmatique et syntagmatique. Nous avons réalisé un outil, inspiré de [RIC'98], qui affiche les similarités textuelles inter-feuillet par calcul de matrices de similarités. Voici comment il fonctionne : supposons que l'on veuille comparer le fragment de texte « les oiseaux chantent sur les arbres » et « l'oiseau gazouille sur l'arbre ». L'apostrophe de la seconde phrase est remplacée par un espace pour le traitement. La matrice de distance suivante est calculée :

	les	oiseaux	chantent	sur	les	arbres
L	0,667	1	1	1	0,667	1
oiseau	0,833	0,143	0,875	0,833	0,833	1
gazouille	0,889	1	1	0,889	0,889	0,889
Sur	1	0,714	1	0	1	0,833
L	0,667	1	1	1	0,667	1
Arbre	1	1	0,875	0,8	1	0,167

La distance entre le mot figurant sur une ligne et sur une colonne peut se lire dans la cellule à l'intersection de cette ligne et colonne. À partir de cette table, une image est calculée (Fig. 4).

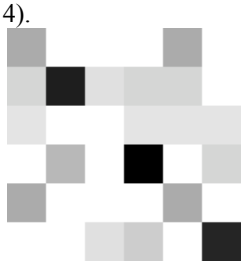

Fig. 4 : Concordance visuelle entre les deux phrases « les oiseaux chantent sur les arbres » et « l'oiseau gazouille sur l'arbre », une distance nulle étant représentée en noire et une distance maximale en blanc.

Une diagonale de points noirs apparaît visuellement, bien que les mots ne soient pas toujours identiques. En appliquant cette technique à tous les feuillets d'un corpus, on peut voir apparaître des motifs diagonaux qui indiquent une corrélation entre des feuillets deux à deux (Figure 4). Cette corrélation implique une réécriture ou une forte similarité. Appliqué à plusieurs feuillets, cette méthode permet de voir immédiatement les corrélations entre feuillets. La méthode est résistante aux fautes d'orthographes et aux formes fléchies des mots, contrairement aux programmes de différenciations textuelles.

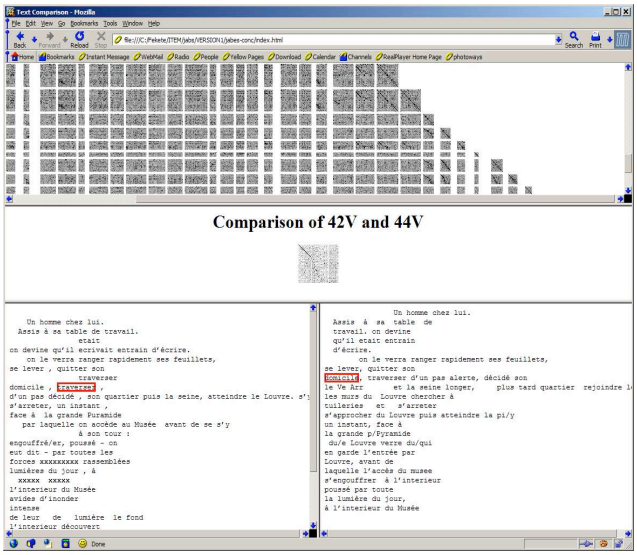

Fig. 5 : Résultat de l'application de la concordance visuelle sur 20 feuillets d'un manuscrit de Jabès et affichage de deux feuillets synchronisés avec une forte corrélation au début des feuillets.

4.3 Visualisation et navigation dans un dossier génétique

Une fois constitué, un dossier génétique contient plusieurs types de manuscrits : notes, brouillons, feuillets dactylographiés et éventuellement bon à tirer. Ces feuillets peuvent être structurés et décrits à l'aide de TEI et transformés en représentations graphiques facilitant la navigation (Fig. 6).

5 Conclusion et perspectives

Dans cet article, nous avons décrit un codage de manuscrit sur trois niveaux : l'image, la description scripto-graphique et la description du dossier. Nous avons aussi montré comment ce codage peut être exploité à l'aide d'outils facilitant la lecture, l'analyse et l'exploration de documents de genèse.

Nous pensons que ce codage est suffisamment flexible pour s'adapter à tout type de manuscrit : historique stabilisé, brouillons littéraires, cahier de laboratoire, etc. Contrairement aux autres documents numériques connus, un seul niveau de description ne suffit pas pour capter la structure d'un manuscrit. Une fois cette réalité admise, des outils informatiques peuvent être envisagés pour développer le champ d'étude des manuscrits qui reste encore dépendant d'outils du marché mal adaptés.

Notre codage est actuellement réalisé manuellement mais nous travaillons parallèlement à des outils de transcription semi-automatique pour faciliter la saisie et la mise en relation des couches entre elles. Nous mettons en place également une approche collaborative pour la constitution d'un corpus de transcriptions. Nous pensons que ces approches de codage, structuration et annotations collaboratives nous permettront enfin d'avancer efficacement dans l'exploitation en ligne des manuscrits.

Nous espérons que la communauté de la reconnaissance de formes et du traitement de l'image pourra nous aider à résoudre des problèmes complexes, comme l'appariement automatique transcription/image, l'extraction automatique ou la correction de traits ainsi que la reconnaissance globale de mots à partir de leur image ou de leur ductus, entre autres. L'extraction automatique de « zones » du document constituerait une avancée importante pour l'étude du manuscrit notamment pour la prise en compte des granularités variables.

Références

[BOZ'97] Bozzi, A. and S. Calabretto (1997). *The Digital Library and Computational Philology: The BAMBI Project*. Research and Advanced Technology for Digital Libraries. First European Conference, ECDL '97, Pisa, Italy, Springer-Verlag.

[ESP'99] Esprit Project (1999). MASTER: Manuscript Access through Standards for Electronic Records.

[FEK'98] Fekete, J.-D. (1998). *Expérience de codage de document à intérêt graphique à l'aide de TEI*. Actes du congrès Eurotex 98, Saint Malo.

[FEK'99] Fekete, J.-D. and N. Dufournaud (1999). "Analyse historique de sources manuscrites : application de TEI à un corpus de lettres de rémission du XVIième siècle." *Numéro spécial "Numérisation et structuration des documents anciens" de la revue "Document Numérique"* **3**(1-2): 117-134.

[FER'03] Ferraiolo, J., J. Fujisawa, et al. (2003). Scalable Vector Graphics (SVG) 1.1 Specification. *W3C Recommendation*.

[GUS'99] Gusnard de Ventabert (nom collectif) (1999). "Représentation et exploitation électronique des documents anciens numérisés." *Numéro spécial "Numérisation et structuration des documents anciens" de la revue "Document Numérique"* **3**(1-2): 57-73.

[LEC'98] Lecolinet, E., L. Likforman-Sulem, et al. (1998). *An integrated reading and editing environment for scholarly research on literary works and their handwritten sources*. Proceedings of the third ACM conference on Digital libraries, Pittsburgh, Pennsylvania, United States, ACM Press New York, NY, USA.

[RIC'98] Richy, H. and J. André (1998). *Édition comparative et hypertextuelle*. Document Électronique (Actes du Colloque International sur le), Rabat.

[SPE'02] Sperberg-McQueen, C. M. and L. Burnard (2002). *TEI P4: Guidelines for Electronic Text Encoding and Interchange*, Text Encoding Initiative Consortium (Oxford, Providence, Charlottesville, Bergen).

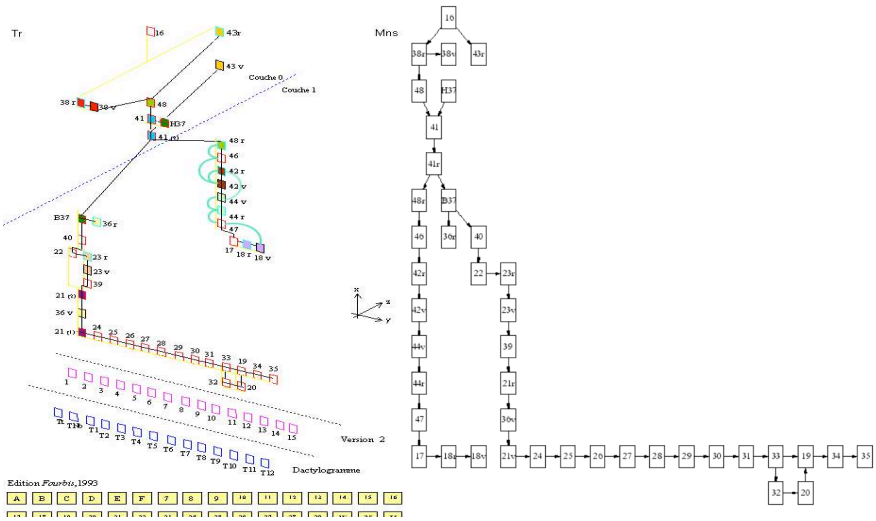

Fig. 6 : A gauche, représentation des axes dessinée à la main. A droite, représentation générée automatiquement et liée aux feuillets manuscrits, permettant aussi de vérifier la cohérence des déductions du chercheur sur l'ordre des feuillets.