



HAL
open science

Du corpus à la région : représentations multi-échelles pour les manuscrits modernes

Aurèle Crasson, Jean-Daniel Fekete

► **To cite this version:**

Aurèle Crasson, Jean-Daniel Fekete. Du corpus à la région : représentations multi-échelles pour les manuscrits modernes. *Linguistica Computazionale*, 2004, XX-XXI, pp.111-128. hal-00875912

HAL Id: hal-00875912

<https://inria.hal.science/hal-00875912v1>

Submitted on 5 Aug 2019

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

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

DU CORPUS A LA RÉGION
Représentations multi échelles pour les manuscrits modernes

CRASSON AURÈLE*, JEAN-DANIEL FEKETE**

*ITEM CNRS

Aurele.Crasson@ens.fr,

** INRIA Futurs/LR

Jean-Daniel.Fekete@inria.fr

Abstract – Beyond the manuscripts sources and related documents, there are processes of work that imply representations for memorization and analysis. We propose some tools to help interpretate and navigate into the manuscripts information. These are made for representing manuscript corpora and the relationships between various data and types of unit that characterize this kind of document.

Keywords – manuscripts, genetic studies, representation, encoding, tools.

1. TERRITOIRES, REGIONS, FRAGMENTS : LES TRAVERSES DU
MANUSCRIT D'AUTEUR

Le présent travail s'intéresse aux systèmes de représentation de documents hybrides dans une perspective d'édition électronique savante. L'étude porte sur les manuscrits littéraires, dans lesquels dessins, graphies, textes et supports s'assemblent, se superposent et posent, de part la nature de leurs informations, un problème de codage et de représentation.

L'objectif consiste à transposer ce support matériel « d'objet manuscrit » dans un environnement informatisé et de développer des dispositifs de représentation adéquats permettant aux chercheurs d'appréhender les données selon différents points de vue.

1.1 Qu'est-ce que le manuscrit ?

Le manuscrit est un objet qui s'inscrit dans l'histoire ; productions, corrections, latences, éditions, changements de main, autant d'empreintes temporelles d'une activité processuelle auxquelles se confrontent tant un lecteur averti que le critique.

D'un point de vue visuel, lisible et sensible, le brouillon d'auteur est constitué d'un ensemble d'informations hybrides qui, outre textes et dessins, renferme des signes métalangagiers et idiosyncrasiques qui s'informent mutuellement.

Il met en présence, sur un support d'inscription, des traces figées d'une élaboration en cours, d'un travail visant un devenir d'œuvre.

Qu'on le décrive en tant qu'image, que l'on transcrive ses textes ou que l'on s'astreigne à un repérage exhaustif des données de ce support, il y a nécessairement perte ou prolifération d'informations qui aboutissent à rendre l'objet plus complexe ou trop simplifié. Une appréhension du document par des niveaux tels que les données graphiques, les coordonnées des blocs d'informations dans le support, les opérations d'écriture, les repérages temporels, etc., permettrait de circonscrire des recherches particulières en se détachant du réseau sous-jacent qui anime et contextualise tous types d'informations.

Quand on le regarde comme un support qui a servi à fixer les traces d'une langue particulière tout en donnant l'illusion d'un mouvement qui se prolonge, d'une respiration, ce sont les questions relatives aux usages de l'écriture tant du côté de la scription¹ que de celui de l'extériorisation et de l'inscription de la langue² qui interpellent spontanément.

L'analyse du « fait d'écriture » et l'apparence du manuscrit renvoient à un autre type d'information : les chronologies d'écritures. On les représente généralement selon des axes

¹ Usages de l'espace topographique, transcription des codes typographiques, inventions idiosyncrasiques, traitement des corrections, etc.

² Construction du texte, lexique, syntaxe, sens, etc.

paradigmatique³ et syntagmatique. Si des empreintes d'effacement ou des traits de biffure par exemple, à travers les ruptures de gestes de scription ou dans les interruptions de la flèche narrative sont repérables, il n'existe pas réellement de preuves tangibles qu'à un temps T tel mot ait été ajouté par rapport au temps T' du reste de la phrase.

Ratures, surcharges, suppressions, substitutions ou encore déplacements constituent des indices de variations chronologiques à une échelle micro-génétique ; les interruptions de raisonnement comme de toutes suites logiques, se mesurent à une échelle macro-génétique. L'interprétation est cependant extrêmement subjective. L'observation du manuscrit se fait dans le rapprochement de ces deux types de chronologies (focalisation – défocalisation) qui permet de comparer les périodes de productions de celles des relectures et corrections.

1.2 Représentation informatisée

Le déplacement sur support numérique du document manuscrit pose d'emblée la question de l'unité élémentaire et de ses paradigmes. Cette question constitue pour quiconque cherche à représenter ce document un des problèmes majeurs. La spécificité et la richesse du manuscrit sont telles que n'importe quel grain le constituant pourrait faire office d'unité référentielle.

Dans la pratique, c'est généralement le feuillet, ou toute autre origine matérielle unitaire qui la plupart du temps fait office d'unité de référence.

Cette unité a fonctionné jusqu'à l'arrivée des ordinateurs parce qu'il était difficile, sauf à l'annoter ou le découper, de descendre au-dessous du feuillet.

Transposés tels quel dans un environnement informatique, l'unité feuillet constitue une représentation très en deçà des potentialités du numérique et notamment de l'hypertexte.

³ Le paradigme est pris ici avec un déplacement par rapport à la façon dont il est utilisé en linguistique. D'un point de vue génétique, il n'est pas possible d'éliminer le temporel dans les empilements de réécriture.

Il n'y a pas lieu à nos yeux de fixer une unité représentative du manuscrit parmi les multiples possibilités que présente ce support⁴, dans la mesure où chacune d'elle renvoie à du singulier. Il s'agit au contraire de multiplier non seulement les choix de nombre, de taille, de dimension mais aussi de détermination d'une unité (détermination selon leurs traces topographiques, selon des logiques de recherche, selon des logiques d'écriture, etc.)

Ce principe, qui entraîne un certain détachement par rapport au document palpable, n'exclut pas pour autant sa référence comme une des possibilités de représentation. Il permet seulement de multiplier les approches d'un même document et, de fait, les possibilités de son interprétation.

2. CONCEPTS DE REPRESENTATION

Le titre de cet article sous-entend la recherche de systèmes de représentation et d'aide à l'interprétation du manuscrit, support de travail du généticien. Il ne s'agit donc pas ici de présenter une édition ni même un concept d'édition.

Notre démarche consiste à concevoir des dispositifs permettant de donner à voir sur un support numérique un ensemble d'objets matériels transposés selon des logiques spécifiques à ce support.

Ce travail, dont l'objectif est d'aider le lecteur comme le chercheur à manipuler des documents complexes, consistera à concevoir des interfaces permettant d'isoler des nœuds d'informations, de composer des séquences ou de restituer symboliquement le feuillet à l'écran.

L'édition génétique électronique est encore aujourd'hui équivoque, le mot même est souvent employé dès lors qu'un ensemble de documents stockés sur un ordinateur sont consultables. Certains chercheurs-éditeurs pensent que les bases de données, le PDF ou l'HTML offrent sans trop d'effort un moyen satisfaisant d'organiser des corpus et d'élaborer ainsi une publication critique qui sera souvent sous forme imprimée. Loin d'être inintéressante,

⁴ Une unité est toujours définie par son contexte immédiat : une zone du feuillet, le feuillet par rapport à une partie du corpus, le corpus comme partie de l'œuvre, autant de choix possibles à toutes les échelles du document.

cette démarche se heurte malgré tout à l'évidente inadéquation des modèles de programme « grand public » au support manuscrit. De plus, considérant que l'établissement du corpus est une démarche individuelle qui vise ses seules analyses, le chercheur empêche toute forme de transmission à d'autres travaillant sur les mêmes données.

2.1. Régions et réseaux

Le numérique autorise la modélisation d'informations singulières. En l'occurrence, il permettrait pour le manuscrit, de stabiliser les règles communes de l'écriture, des universaux du manuscrit d'auteur, de repérer, et définir comme tels, les signes idiosyncrasiques ou graphiques ainsi que les éléments méthodologiques de la discipline génétique. En outre, il permettrait de rendre visible les relations qui unissent toutes ces informations.

Afin de bien les rendre utilisables en tant qu'unités individuelles, ces informations doivent alors être décrites de façon à pouvoir être facilement encodées lors de l'établissement du document. Ces informations discrétisées, peuvent alors être commutées au gré des diverses logiques de raisonnement et d'analyse ; ces connexions pouvant elles-mêmes se définir en tant que lien-type décrivant la qualité de la relation.

Enfin, pour être visualisées selon des organisations propres à des logiques de recherche et à des points de vue macro ou microstructurels, il s'agit de donner à l'utilisateur, les moyens de choisir et modifier la taille de l'unité de référence – mot, zones visuellement isolées, unité matérielle (page), blancs (surfaces entières ou frontières des zones couvertes).

L'unité géographique « région » (Figure 1) a été ainsi définie.

Cette référence se fonde, comme nous l'avons vu, selon des logiques d'analyses propres à la démarche de recherche (telle entité logique plutôt que telle autre) ou selon les contours que la trace scripto-graphique a virtuellement formés.

Il s'agit d'un « contenant » non arbitraire, qui inclut un fragment textuel ou autre, lequel se lie à d'autres informations. Des fragments sont associés à des régions. Les régions sont topographiques ; les fragments, textuels ou graphiques.

Ceci permet d'envisager le manuscrit comme un ensemble de régions structurées par un réseau latent, chacune d'elles étant constituée en tant que nœud potentiel à partir duquel ou vers lequel se fixent des liens. Ces régions élémentaires et ouvertes, ne sont pas arbitraires dans la mesure où chaque document extériorise son type d'unité : il s'agit de trouver une logique qui caractérise le corpus en accentuant sa singularité afin d'élaborer des représentations constituant une aide réelle à la mémorisation des dossiers de genèse.

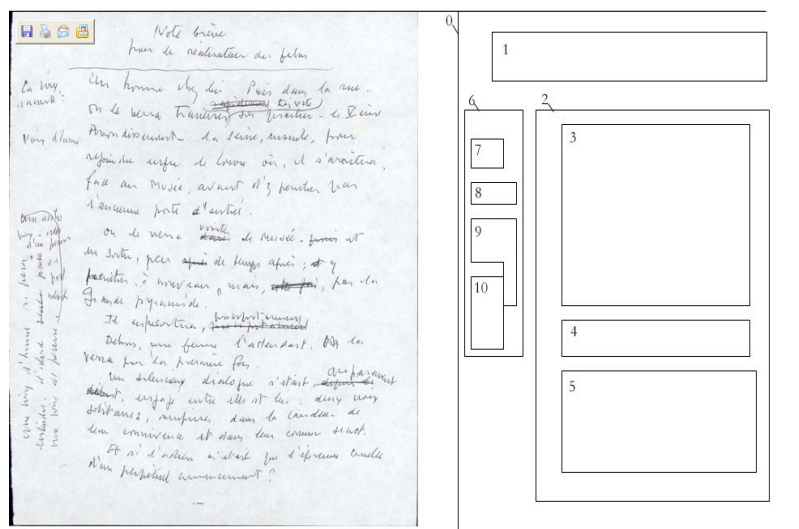


Figure 1 : Découpage du manuscrit en régions

L'étude pratique des avant-textes faite à partir de photocopies entraîne de facto des dispositifs compliqués de mémorisation : annotations, signaux, balisages d'événements, mais aussi post-it, découpages, compilations multiples, etc. Avec l'informatique, et grâce à des formes plus abstraites, ces notes, indications graphiques ou procédés mentaux de mémorisation, tout type d'ajout d'informations, engendrent des dispositifs cognitifs nouveaux mieux adaptés à un support riche d'informations.

2.2. Outils et développements

Lors du colloque sur les disciplines philologiques et les technologies numériques qui s'est tenu au Ciocco du 6 au 11 septembre 2003, nous avons pris le parti de réduire les enjeux de l'édition à la présentation de certains outils de visualisation et d'analyse de données manuscrites.

Nous nous sommes ainsi intéressés à trois niveaux d'appréhension des données :

- 1) le dossier génétique
- 2) les variantes dans les transcriptions
- 3) la lecture en ligne de manuscrit

Ces trois niveaux font partie d'un ensemble plus important d'outils que nous concevons et réalisons pour l'aide à l'analyse et à l'annotation en général.

À travers l'étude d'un corpus⁵ ayant occupé deux ans de travail, nous avons cherché à reproduire informatiquement le graphe de représentation de la genèse - dessiné, à l'époque à la main et dont nous reproduisons ici l'image (Figure 2) -, à partir de la description des informations génétiques (Figure 3).

Ce graphique s'est avéré être un outil très utile pour localiser les feuillets dans la masse complexe des brouillons, même ordonnés, et pour évaluer rapidement le résultat du classement génétique voire même les étapes chronologiques entre chaque période d'écriture.

⁵ *Cela a eu lieu*, manuscrit d'Edmond Jabès

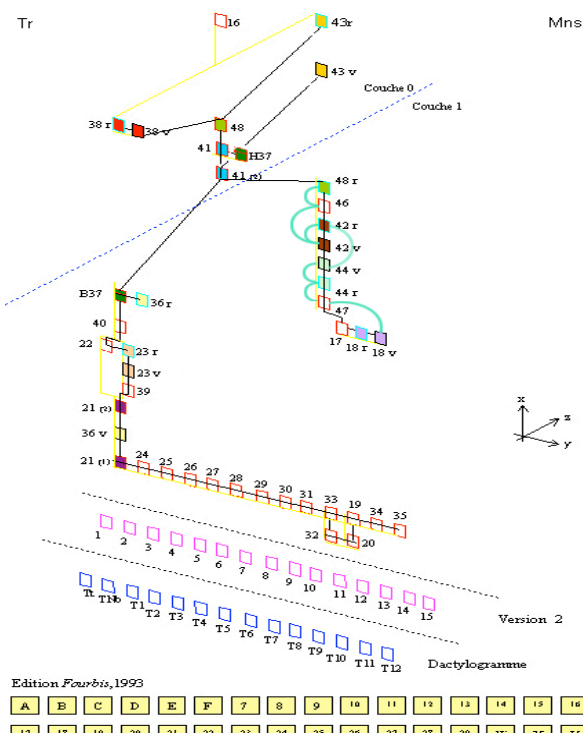


Figure 2 : Représentation tridimensionnelle du dossier "Cela a eu lieu" d'Edmond Jabès

Un outil de concordance approximative (Figure 6), a été développé pour l'analyse des feuillets transcrits⁶, autorisant la comparaison lexicale des feuillets et leurs potentialités de rapprochement.

Enfin, l'approche d'une entité précise du manuscrit nécessitant une permanente connexion au contexte global - d'une région spécifique ou d'une quelconque zone hiérarchiquement supérieure (le feuillet, le dossier génétique, l'œuvre, etc.) -, nous avons également élaboré un dispositif permettant de focaliser une zone du manuscrit tout en la superposant à son contexte. (Figure 7)

⁶ Nous n'évoquons pas ici les problèmes des transcriptions, provenant notamment d'une lacune concernant les conventions de transcription et du manque d'outil permettant de saisir « à l'identique » les blocs graphiques du manuscrit.

1) Le dossier génétique

Un dossier génétique est constitué de feuillets manuscrits et dactylographiés. En cas d'absence de pagination autographe, ces feuillets doivent être en priorité numérotés et organisés pour en faciliter l'analyse. La genèse se représente, selon un principe adopté par les linguistes, par des axes dans lesquels l'unité de référence adoptée est représentée. Dans le cas particulier du dossier Jabès un troisième axe d'incertitudes a été créé pour représenter des données inclassables⁷ sur les autres axes temporels, syntagmatique et paradigmatic.

Pour faciliter la compréhension de l'analyse et vérifier automatiquement la plausibilité des hypothèses de classements chronologiques, nous avons construit la représentation de la figure 2.

Cette représentation utilise l'axe Y pour le syntagmatique, l'axe X pour le paradigmatic et l'axe Z pour représenter l'incertitude. La branche la plus éloignée en Z étant une branche morte du récit.

Cette représentation construite initialement à la main présente clairement les relations entre les feuillets. Il faut bien rappeler qu'il s'agit d'une vue simplifiée du dossier génétique : des relations existent au sein d'un même feuillet.

Inspirés par ce graphe, nous avons réalisé un outil qui construit automatiquement une représentation 2D d'un dossier à partir d'une description textuelle des relations entre les feuillets (Figure 3).

⁷ Compte tenu de la fréquence des incertitudes dans les avant-textes, ce principe peut tout à fait être réitéré à différentes échelles de régions.

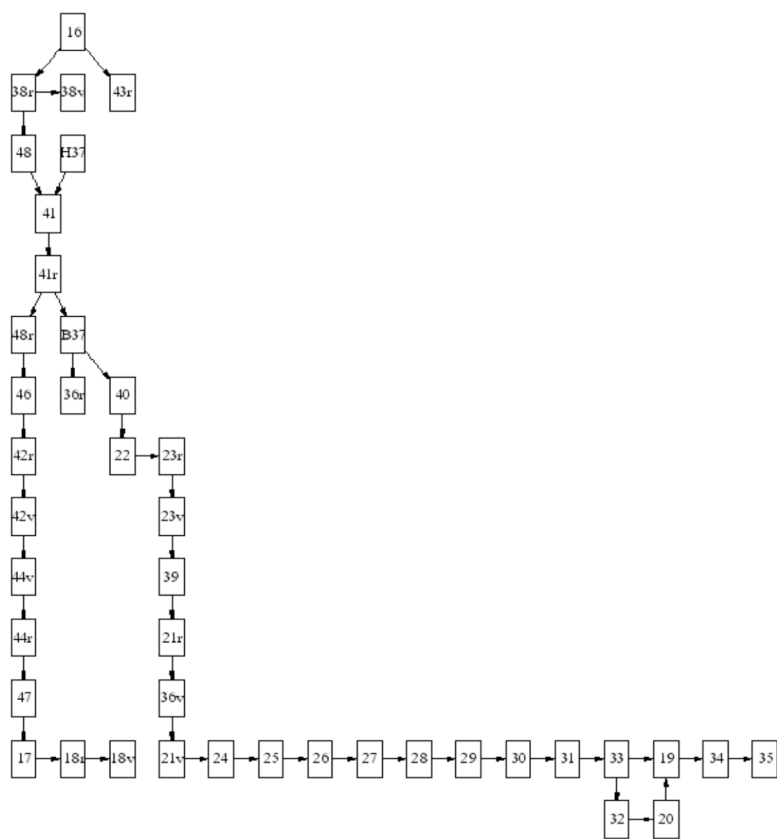


Figure 3 : Représentation bidimensionnelle du dossier génétique généré automatiquement à partir d'une description des relations entre feuillets.

Cette description a la forme suivante :

para "16" -> "38r";

synta "38r" -> "38v";

para "38r" -> "48" -> "41" -> "41r";

Qui peut se lire ainsi : le feuillet 16 est suivi du feuillet 38r sur l'axe paradigmatique. Le feuillet 38r est suivi du feuillet 38v sur l'axe syntagmatique, etc.

Cette représentation textuelle (qui pourrait être saisie graphiquement) permet de générer automatiquement et instantanément la représentation graphique, facilitant la vérification des hypothèses génétiques sur le dossier. Une erreur produirait des cycles curieux ou des rattachements inattendus, immédiatement visibles. La représentation est visible à partir d'un navigateur Web ; les nœuds étant, par ailleurs, liés aux pages de manuscrits l'exploration à l'échelle du dossier et l'accès aux feuillets s'en trouve facilité.

2) Variantes dans les transcriptions

La construction de la structure du dossier génétique est un travail long et difficile. Pour faciliter la lecture du manuscrit, chaque feuillet est transcrit selon le principe théorique d'une interprétation minimale. Grâce à un balayage visuel entre cette transcription et le fac-similé, les différents indices nécessaires à l'ordonnement chronologique du texte de l'œuvre peuvent être repérés et fixés pour d'éventuelles interprétations.

Pour faciliter la recherche des segments communs entre les multiples versions du récit, nous avons réalisé un outil d'analyse visuelle de concordance entre documents textuels. La recherche de concordances est un travail classique d'analyse textuelle, mais il se limite généralement à rechercher les occurrences d'un mot ou d'une phrase précise tout au long d'un document ou d'un corpus. Pour l'étude génétique, il est indispensable de connaître tous les segments de texte qui pourraient être des réécritures d'autres segments placés sur des feuillets différents. Parce qu'il s'agit de textes en processus, la similarité entre feuillet peut être imprécise. Ainsi, contrairement à la recherche de concordance classique, la

recherche ne se situe pas au niveau d'une phrase dans un corpus mais d'une transcription de feuillet à d'autres feuillets. Certains mots peuvent avoir été altérés, d'autres insérés, supprimés ou déplacés, etc.

Nous avons réalisé un système d'analyse graphique basé sur des matrices de similarité entre textes. 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, l'image suivante est calculée :

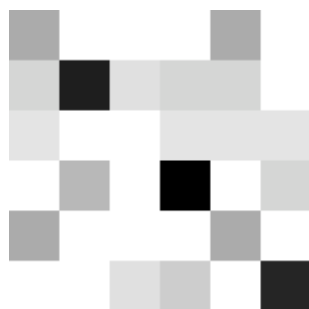


Figure 4 : Concordance visuelle entre les deux phrases « les oiseaux chantent sur les arbres » et « l'oiseau gazouille sur l'arbre », une distance nulle étant représentée en noire et une distance maximale en blanc.

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 5). Cette corrélation implique une réécriture ou une forte similarité.

17 18R 18V 19 20 21(1) 21(2) 22 23R 23V 24 25 26 27 28 29 30 31 32 33 34 35 36R 36V 37 38R 38V 39 40 41(1) 41(2) 42R 42V 43R 43V 44R 44V 45R 45V 46 47 48R 48V

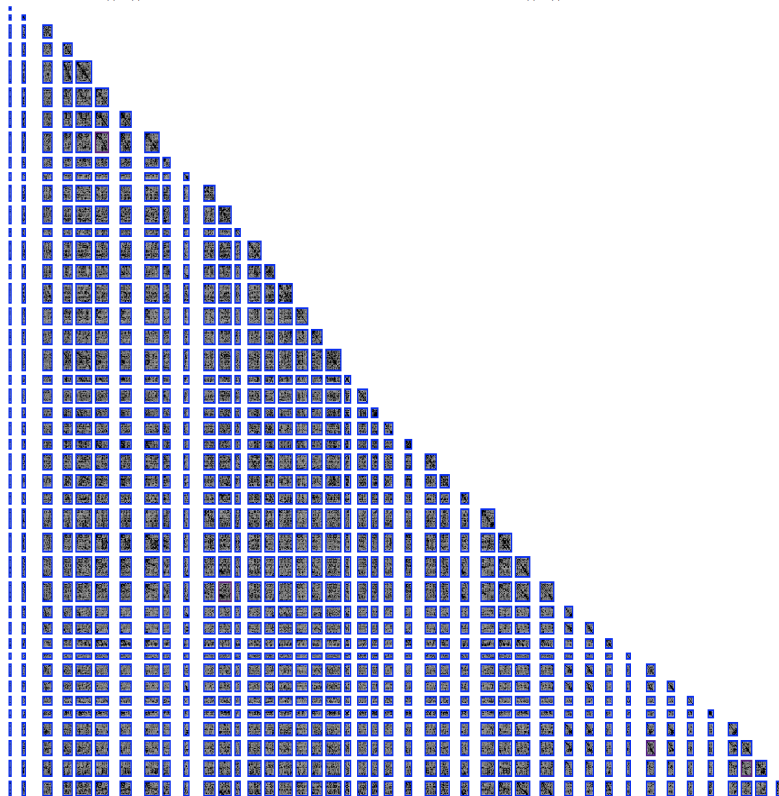


Figure 5 : Calcul systématique des corrélations entre feuillets des manuscrits d'Edmond Jabès transcrits.

L'utilisation d'une distance inter-mots permet d'adapter aux variantes orthographiques et grammaticales (singulier/pluriel par exemple) ou à des fautes d'orthographe. Une telle représentation facilite donc la structuration du dossier génétique.

À partir du calcul systématique des corrélations, notre outil permet de comparer plus précisément les similarités en cliquant sur une

des images miniatures pour comparer les deux transcriptions (Figure 6).

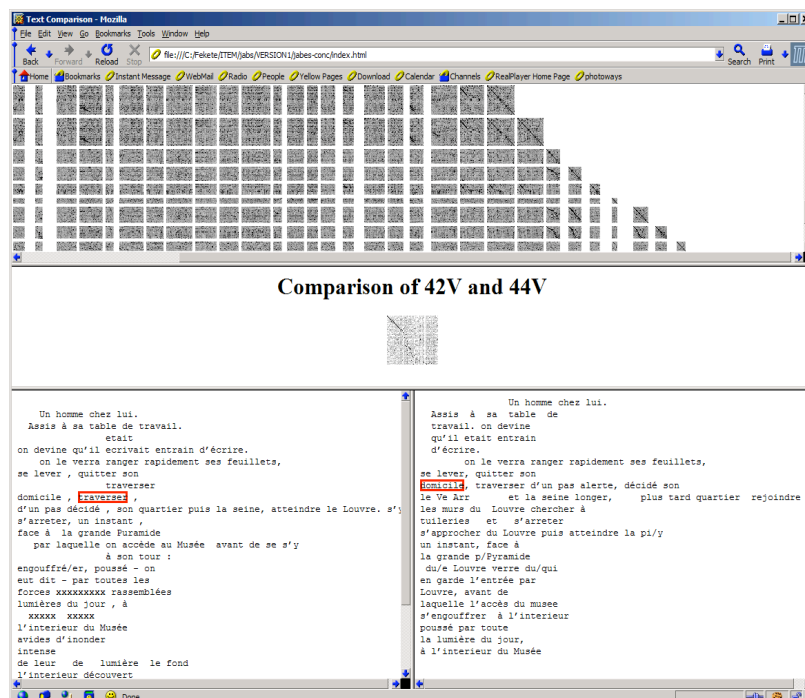


Figure 6 : comparaison détaillée de deux textes visiblement similaires (<http://www.lri.fr/~fekete/concordance>).

3) Lecture en ligne

La structuration d'un dossier génétique repose sur une analyse interprétative qui commence dès la transcription. Menée de manière totalement informatique en ligne, la transcription est particulièrement pénible car les outils d'aide sont pratiquement inexistants. Les programmes d'édition d'images classiques offrent une loupe et des touches permettent de naviguer de gauche à droite et de haut en bas. La loupe est indispensable mais elle a le défaut de cacher le contexte global de la page, masquant totalement les

parties de l'image en dehors de la partie zoomée. Notre système offre à la fois un zoom et une intégration continue entre la partie zoomée et le reste du document, facilitant une vue globale et locale et même temps. Pratiquement, la vue globale évite d'oublier un détail périphérique à la zone textuelle à transcrire.

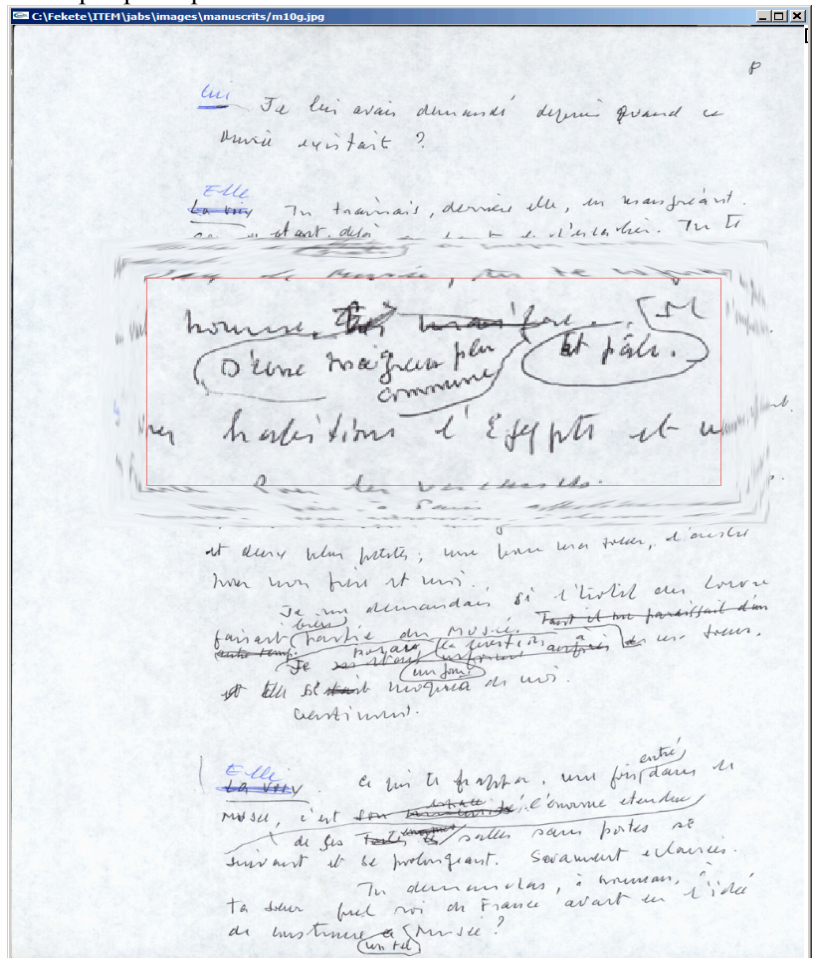


Figure 7 : Représentation "fisheye" d'une page de manuscrit de Jabès, montrant à la fois l'ensemble de la page à peine lisible et un détail agrandi. La zone focale peut être fixée et déplacée interactivement (<http://www.lri.fr/~fekete/manuscript>).

3. TRAVAUX EN COURS ET PERSPECTIVES

Les outils présentés dans ce colloque montrent,

- d'une part, que l'on mémorise mieux des dossiers contenant des informations complexes à travers des représentations abstraites. En optant pour une unité⁸ de référence, il est alors possible de montrer les relations entre ces unités et d'affirmer ce principe de navigation sur de plus fines. Cela reste valable pour tous les niveaux d'information – y compris le mot.
- D'autre part, on peut affirmer que ces représentations peuvent être générées automatiquement à partir d'un encodage stabilisé.

Ces outils pour l'heure ne sont pas intégrés dans une plateforme figée d'aide à la transcription et à l'analyse génétique textuelle. Nous sommes réticents à l'idée d'intégration et de plateforme « couteau Suisse » permettant de scanner, traiter les images, transcrire, structurer, éditer etc.

Notre vision est plus modulaire et flexible. Nous pensons que la création d'outils spécifiques doit être simplifiée et encouragée. Le point clé pour nous est la mise en place d'une infrastructure globale basée sur des technologies WEB déjà existantes et orientées vers le travail collaboratif.

Les outils et représentations décrits ici ne sont que des exemples relativement simples d'un ensemble d'outils facilitant les analyses de génétique textuelle. Ces analyses sont multiples, à la fois en nature mais aussi en termes de problématiques spécifiques aux chercheurs ou aux auteurs. C'est pourquoi une architecture monolithique nous paraît inadaptée.

Parmi les modules sur lesquels notre travail se porte actuellement, le plus important est certainement l'aide à la transcription en ligne. Plusieurs travaux ont déjà été réalisés dans des projets de recherche⁹ pour faciliter la transcription, mais il s'agissait principalement de manuscrits historiques relativement réguliers,

⁸ Dans le cas particulier du dossier Jabès la région s'est définie sur la base du feuillet, elle a ensuite été décomposée pour la fabrication de parcours génétiques, en zones visuellement discrétisables.

⁹ Andrea Bozzi, Sylvie Calabretto, the BAMBI Project.

très proches de textes imprimés. Les manuscrits littéraires à la fois plus complexes, plus graphiques et atomisés nécessitent un outil spécifique pour passer d'une part, de l'image à une transcription diplomatique précise, d'autre part, à l'étape de structuration temporelle, cœur même du travail de génétique textuelle.

Contrairement à la plupart des outils réalisés jusqu'à présent qui étaient dédiés à la création d'un « produit final », notre préoccupation concerne des outils d'établissement d'un corpus scientifique qui articulent des « processus de travail » dans une visée d'édition scientifique. Ce changement de perspective nous semble fondamental du point de vue de la création de savoirs et de l'aide apportée par l'outil informatique qui rend ainsi visibles, réutilisables ou améliorables, les étapes intermédiaires d'analyse et de recherche et, dans tous les cas, pédagogiquement utiles.

La production traditionnelle d'articles ou de monographies devient alors un cas particulier d'un ensemble presque illimité de productions tels des hypertextes ou plus généralement des espaces de navigation virtuelle, des descriptions, classifications, commentaires critiques, présentations pédagogiques, etc., - liste limitée par notre seule imagination.

Ainsi, à partir d'une bonne structuration des manuscrits littéraires, nous pourrions proposer de nouvelles et multiples présentations, visualisations ou animations explicitant sinon le processus entier d'une création littéraire du moins certaines étapes significatives. S'il y a urgence de développer en premier lieu des outils pour l'analyse et les représentations des manuscrits, l'édition d'un travail de genèse en passe aussi obligatoirement par une réflexion d'ordre plus générale sur l'adaptation du support manuscrit au numérique.

RÉFÉRENCES

- BOZZI A., CALABRETTO S., *The Digital Library and Computational Philology : The BAMBI Project*. ECDL 1997, 269-285., 1997
- de BIASI P.-M., *La génétique des textes*, Nathan Université, Paris, 2000.
- CERQUIGLINI, B., *Éloge de la variante. Histoire critique de la philologie*, Seuil, Paris, 1989
- CONTAT M., FERRER D. (éd.), *Pourquoi la critique génétique ? Méthodes, théories*, CNRS éditions, Paris, 1998.
- CRASSON A., *Pour une lecture ouverte du texte électronique*, Making Texts for the Next Century, Literary&linguistic Computing , éd. Oxford University Press, vol. 15, N°1, Oxford , 2000
- CRASSON A., *L'œuvre d'Edmond Jabès peut-elle se lire sous forme de 0 et de 1 ?*, Portrait d'Edmond Jabès, éd. Bibliothèque Nationale de France, Paris, 2000
- FEKETE J.-D., *Expérience de codage de document à intérêt graphique à l'aide de TEI*. Actes du congrès Eurotex 98, Saint Malo., 1998
- FEKETE J.-D., DUFOURNAUD N., *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. Paris, 1999
- GRESILLON A.(1994), *Eléments de critique génétique*, PUF, Paris, 1994.
- HAY L., *la littérature des écrivains*. Etudes de critique génétique, Corti, Paris, 2002.
- ANIS J., LEBRAVE J.-L.(éds), *Texte et ordinateur : les mutations du lire-écrire*, Ed. de l'Espace européen 1991 ; deuxième édition augmentée, CRL Université de Paris X-Nanterre, Paris, 1993.
- LECOLINET E., LIKFORMAN-SULEM L., *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, 1998
- LECOLINET E., ROBERT L., *Conception d'un poste d'édition et de lecture d'hypermédias littéraires*. Numéro spécial "Numérisation et structuration des documents anciens" de la revue "Document Numérique" 3(1-2): 103-115, 1999