



**Genesis**

Manuscrits – Recherche – Invention

**36 | 2013**

**Proust, 1913**

---

## Le codage en TEI des brouillons de Proust : vers l'édition numérique

Julie André et Elena Pierazzo

---



### Édition électronique

URL : <http://journals.openedition.org/genesis/1159>

DOI : 10.4000/genesis.1159

ISSN : 2268-1590

### Éditeur :

Presses universitaires de Paris Sorbonne (PUPS), Société internationale de génétique artistique littéraire et scientifique (SIGALES)

### Édition imprimée

Date de publication : 15 juin 2013

Pagination : 155-161

ISBN : 978-2-84050-893-9

ISSN : 1167-5101

### Référence électronique

Julie André et Elena Pierazzo, « Le codage en TEI des brouillons de Proust : vers l'édition numérique », *Genesis* [En ligne], 36 | 2013, mis en ligne le 09 juillet 2015, consulté le 01 mai 2019. URL : <http://journals.openedition.org/genesis/1159> ; DOI : 10.4000/genesis.1159

---

Tous droits réservés

## Le codage en TEI des brouillons de Proust : vers l'édition numérique

Julie André et Elena Pierazzo

*Ce que nous voulons, c'est la littérature qui bouge,  
et saisie dans le moment même où elle semble bouger encore.*

Julien GRACQ, *En lisant, en écrivant*<sup>1</sup>

En général, les éditions de manuscrits adoptent la présentation diplomatique, parfois dite « ultra-diplomatique », qui s'efforce de rendre compte, de manière aussi mimétique que possible, de la disposition spatiale de l'écriture<sup>2</sup>. Bien que ce type d'édition présente de nombreux avantages, une dimension fondamentale lui fait défaut : la dimension processuelle, dynamique. Et si le numérique apparaît souvent comme l'avenir de l'édition, transposer simplement le texte de la page à l'écran se révèle le plus souvent décevant<sup>3</sup>. Les éditions ultra-diplomatiques en ligne sont en général présentées en vis-à-vis du fac-similé du manuscrit, mais cette représentation n'est pas satisfaisante, et cela pour plusieurs raisons : tout d'abord, l'imitation n'est jamais parfaite ; ensuite, c'est à l'utilisateur/lecteur de faire la mise en relation de la transcription avec le document ; enfin, à cause des contraintes spatiales de l'écran, on doit se contenter de présenter une page à la fois, et non, par exemple, une double page – ce qui, dans le cas des cahiers de Proust, trahit la réalité du manuscrit, puisque la double page est, chez Proust, l'espace de l'écriture.

L'ambition des humanités numériques (*Digital Humanities*) consiste précisément à proposer de nouvelles pistes pour présenter et représenter les textes, notamment les textes littéraires. Un outil d'encodage comme la TEI (*Text Encoding Initiative*<sup>4</sup>) mis au point par une équipe internationale de chercheurs a plusieurs avantages : d'une part, la TEI se veut un standard international de codage ; d'autre part, elle peut être progressivement et indéfiniment enrichie par des balises porteuses de

potentialités nouvelles, en particulier en ce qui concerne la représentation de l'écriture au brouillon. C'est grâce à ce système de transcription et à ces nouvelles balises élaborées dans le cadre du « *SIG Manuscripts*<sup>5</sup> » de la TEI que nous avons essayé de proposer une autre représentation de la genèse en construisant à travers quelques folios de l'un des cahiers de brouillon de Proust – le Cahier 46<sup>6</sup> – plusieurs parcours qui reproduisent la dynamique de l'écriture au brouillon et le cheminement tortueux de sa lecture.

1. Julien Gracq, *En lisant, en écrivant*, Paris, José Corti, 1980, p. 284.

2. Cet article a fait l'objet d'une communication sous le titre « Autour d'une séquence et des notes du Cahier 46 : enjeu du codage dans les brouillons de Proust », lors du colloque « Proust, l'œuvre des manuscrits » organisé par Nathalie Mauriac Dyer et Kazuyoshi Yoshikawa à l'ENS et à la BnF les 1<sup>er</sup> et 2 mars 2012.

3. Pour une réflexion sur cette question, voir Elena Pierazzo et Kathryn Sutherland, « The Author's Hand: from Page to screen », dans *Collaborative Research in the Digital Humanities*, dir. Marilyn Deegan et Willard McCarty, Aldershot, Ashgate, 2012, p. 192 sq. ; Françoise Leriche, « Quelle édition pour quel public ? », *Recherches & Travaux*, n° 72, 2008, <<http://recherchestravaux.revues.org/index89.html>>.

4. <[www.tei-c.org](http://www.tei-c.org)>.

5. Ce *Special Interest Group* a élaboré dans les années 2007-2011 une proposition pour le codage des manuscrits modernes et des éditions génétiques. Cette proposition a été réélaborée ensuite par la TEI qui l'a incorporée en décembre 2011. Voir <[www.tei-c.org/SIG/Manuscripts/genetic.html](http://www.tei-c.org/SIG/Manuscripts/genetic.html)>.

6. Le Cahier 46 (NAF 16686) date de 1914-1915. Proust cherche à cette date à introduire le personnage d'Albertine récemment créé dans ce qui deviendra *Le Côté de Guermantes*. Pour une transcription intégrale, voir Julie André, « Le Cahier 46 de Marcel Proust. Transcription et interprétation », thèse de doctorat, Université Paris III, 2009, vol. II.

À travers l’encodage en TEI des brouillons de Proust, il nous importe surtout de rendre sensible une spécificité de l’écriture au brouillon – son aspect dynamique. Pour ce faire, nous nous sommes appuyées sur différents modèles développés par les chercheurs comme ceux de « parcours », de « chemin<sup>7</sup> » ou de diagrammes des unités rédactionnelles<sup>8</sup> pour proposer plusieurs trajets à travers l’entrelacs manuscrit du cahier : les séquences d’écriture et de lecture. Le premier type de séquence reconstitue virtuellement (et en partie hypothétiquement) les étapes de l’écriture d’un passage ; le second entend donner à suivre le chemin que doit parcourir le lecteur pour en lire le dernier état rédactionnel.

Depuis décembre 2011, la TEI offre en effet la possibilité de transcrire les manuscrits selon une approche topologique ou « géographique » fondée sur trois structures ou balises principales : la balise <surface>, qui représente la surface d’écriture à transcrire (telle que la page, la double page ou le folio, etc.), la balise <zone>, qui représente toute division de la page définie par le chercheur, et la balise <line> qui représente tout ce qui est identifiable comme ligne d’écriture. Pour construire nos séquences d’écriture et de lecture, nous avons utilisé l’élément <zone> qui nous a permis de définir des portions de surface écrite indépendantes de la page ou du folio. Puis nous les avons numérotées selon l’ordre chronologique (préssumé) de la rédaction de manière à pouvoir encoder ensuite la dimension processuelle de l’écriture<sup>9</sup> (voir fig. 1, p. 158-159).

La séquence de lecture, quant à elle, reprend cette numérotation mais dans un ordre différent puisqu’elle entend mimer le cheminement de la lecture<sup>10</sup>. Pour lire la séquence de manière continue, nous proposons en effet l’ordre suivant :

Étapes de la séquence de lecture	
a	3.1
b	1
c	6.1
d	6.2
e	2A <sub>1</sub>
f	2A <sub>2</sub>
g	6.3

h	7.1
i	7.2
...	... <sup>11</sup>
m	5

Si le parcours de la séquence de lecture est relativement simple, en revanche, le parcours de la séquence d’écriture (voir fig. 1) repose sur une série d’hypothèses qui visent à reconstituer l’ordre dans lequel ont été écrites les différentes zones composant l’unité textuelle, non pas à l’échelle de la page ou de la double page mais de la séquence choisie. Elle s’avère de ce fait beaucoup plus complexe<sup>12</sup>. Cette reconstitution s’appuie

7. Voir la définition des « chemins génétiques » donnée par Paolo D’Iorio dans « L’infrastructure *Hyper* et son utilisation pour la critique génétique », *Genesis*, n° 27, 2006, p. 181. « Un chemin est un type de contribution établie par un chercheur qui consiste à rassembler d’une façon pertinente une suite chronologique, thématique ou génétique de sources primaires. » Pour un autre exemple, voir également dans l’*HyperNietzsche* ce que les éditeurs appellent « chemin chronologique », « chemin génétique », « chemin thématique ». Inga Gerike, « Les manuscrits et les chemins génétiques du *Voyageur et son ombre* », <[www.hypernietzsche.org/doc/puf/book/hypernietzsche/le-livre.htm](http://www.hypernietzsche.org/doc/puf/book/hypernietzsche/le-livre.htm)>.

8. Pour une démarche similaire dans une édition imprimée, voir les diagrammes des unités textuelles établis par les éditeurs des *Cahier 54*, *Cahier 71*, *Cahier 26*, dans *Cahiers 1 à 75 de la Bibliothèque nationale de France*, dir. Nathalie Mauriac Dyer, Turnhout, BnF/Brepols, 2008-2011, vol. I. Voir aussi N. Mauriac Dyer, « Déchiffrer. Transcrire. Cartographier. Lier. Proust ou le manuscrit apprivoisé », *Genesis*, n° 27, 2006, p. 23 sq.

9. Cependant, il est à noter que lorsque la rédaction n’est pas continue (par exemple, lorsque Proust ajoute les titres « I » et « II » sur les folios 47 et 48 r<sup>os</sup>), la numérotation le précise en distinguant 3.1, 3.2...

10. On rappelle que le parcours de lecture propose ici de lire la séquence achevée à ce stade de la rédaction.

11. Comme nous l’avons mentionné, la séquence se poursuit sur les folios suivants. Il ne faudrait lire la zone numérotée 5 qu’après avoir lu le folio 50 v<sup>o</sup> et suivants.

12. On note d’emblée que la reconstruction de ce parcours concerne principalement les grandes étapes de la rédaction et qu’elle exclut les additions ou notes postérieures à la rédaction principale de la page le plus souvent impossibles à dater. On a distingué les notes de régie pures (notées « nr ») et les ajouts, éventuellement accompagnés de notes de régie (donc mixtes), pour lesquels le premier chiffre indique la zone à laquelle ils se rattachent. Lorsqu’il y a plusieurs ajouts, nous utilisons les lettres pour les distinguer (voir sur le folio 47 r<sup>o</sup> : 1A, 1B) et lorsqu’un ajout se décompose lui-même en différentes parties, nous les avons numérotées : 1B<sub>1</sub>, 1B<sub>2</sub> (voir f<sup>os</sup> 46 v<sup>o</sup>-47 r<sup>o</sup>). Nous repreneons donc ici la notation adoptée pour les diagrammes des unités textuelles dans l’édition des *Cahiers 1 à 75 de la BnF*, éd. citée.

sur les grandes étapes de l'écriture de la séquence<sup>13</sup>, mais bien entendu, on pourrait faire un travail plus approfondi en intégrant la dimension temporelle non seulement à l'échelle de la séquence mais à l'intérieur de chaque zone. Ainsi, on pourrait, par exemple, mettre en évidence à l'intérieur d'une phrase le premier jet, puis la rature et la substitution qui l'a suivie de manière à reconstituer le cheminement de l'écriture à l'échelle de la phrase<sup>14</sup>. Cette reconstitution suppose également toute une réflexion sur la délimitation des unités et sur leur découpage en zones. Par exemple, la zone que nous situons sur les folios 49<sup>ro</sup> et 50<sup>ro</sup> (zone 5) se répartit de part et d'autre de la zone 4 (f° 50<sup>ro</sup>) qui a été écrite avant, et il est difficile sans une lecture attentive préalable de repérer l'existence de deux zones génétiques distinctes sur le folio 49<sup>ro</sup>. L'encodage repose donc sur une interprétation fine de la genèse de la séquence<sup>15</sup>.

Autre difficulté : comment encoder les marginales dans la mesure où l'on sait qu'il existe parfois un continuum entre la note et l'addition<sup>16</sup> dans les brouillons de Proust ? Par exemple, au folio 47<sup>ro</sup> (voir fig. 2), la marginale commence comme une note de régie (« Capitalissime après chemin ») et se poursuit comme une addition (« Comme Aimé... »), elle-même renforcée par un trait de jonction. Faut-il utiliser la balise <metamark> qui sert à encoder les notes de régie ou l'outil <add> qui sert pour les additions ? Là encore, l'encodage lui-même suppose une interprétation.

Pour explorer le potentiel de la nouvelle approche proposée par la TEI, nous avons aussi expérimenté les formes innovantes d'affichage diplomatique que l'encodage de type topologique a rendues possibles. Chaque section correspondant au balisage <zone> peut être pourvue d'un ensemble de coordonnées spatiales, qu'on peut utiliser (par exemple) pour appliquer point par point les zones sur le fac-similé numérique.

Nous avons élaboré une édition prototype, construite autour du fac-similé<sup>17</sup>. Dans ce prototype, l'utilisateur se voit proposer l'image d'une double page du Cahier 46 (voir fig. 3). En cliquant sur l'image, il fait apparaître successivement en surimpression du fac-similé les zones de transcription, dans l'ordre (préssumé) de leur rédaction ou, plus simplement, dans celui de leur dernier

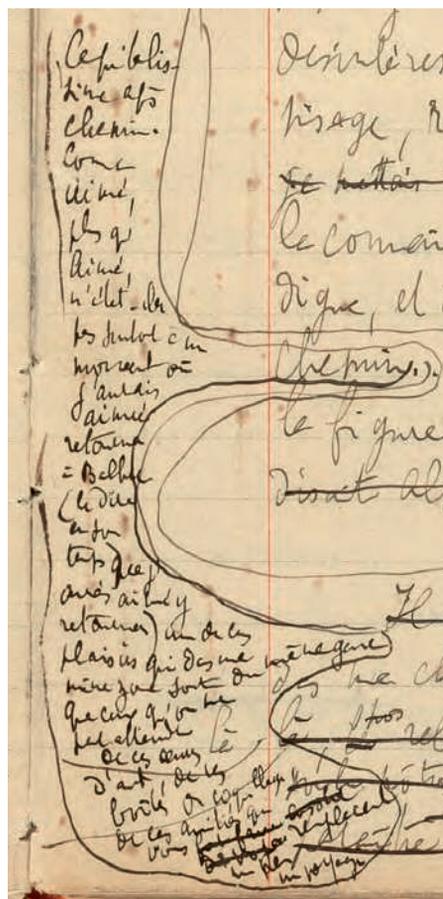


Fig. 2 : Cahier 46, f° 47<sup>ro</sup> (détail ; cf. fig. 1)

13. Il est à noter que la numérotation chronologique (de 1 à 7) est relative : elle s'appuie sur l'extrait de séquence du Cahier 46 envisagé ici. En réalité, la zone numérotée 5 se poursuit sur les folios suivants, de même que la zone 7.

14. Voir Jean-Louis Lebrave, « Du visible au lisible : comment représenter la genèse ? », *Genesis*, n° 27, 2006, p. 11-18.

15. L'ordre de rédaction proposé pour le folio 47<sup>vo</sup> pourrait ainsi surprendre. Notre numérotation repose sur l'hypothèse que la zone 2A<sub>1</sub> a probablement été écrite d'abord (elle répond à la note de régie qui figure en marge du folio 48<sup>ro</sup>). Proust réalise en fait sur le folio 47<sup>vo</sup> le programme indiqué dans la marge du folio 48<sup>ro</sup> et la zone numérotée 6.3 est la reprise du paragraphe barré sur le folio 48<sup>ro</sup>.

16. Voir les articles suivants : Anne Herschberg Pierrot, « Les notes de Proust », *Genesis*, n° 6, « Enjeux critiques », 1994, p. 61-78 et Julie André, « Les notes de Proust, une écriture dialogique », *Genesis*, n° 33, 2011, p. 157-172.

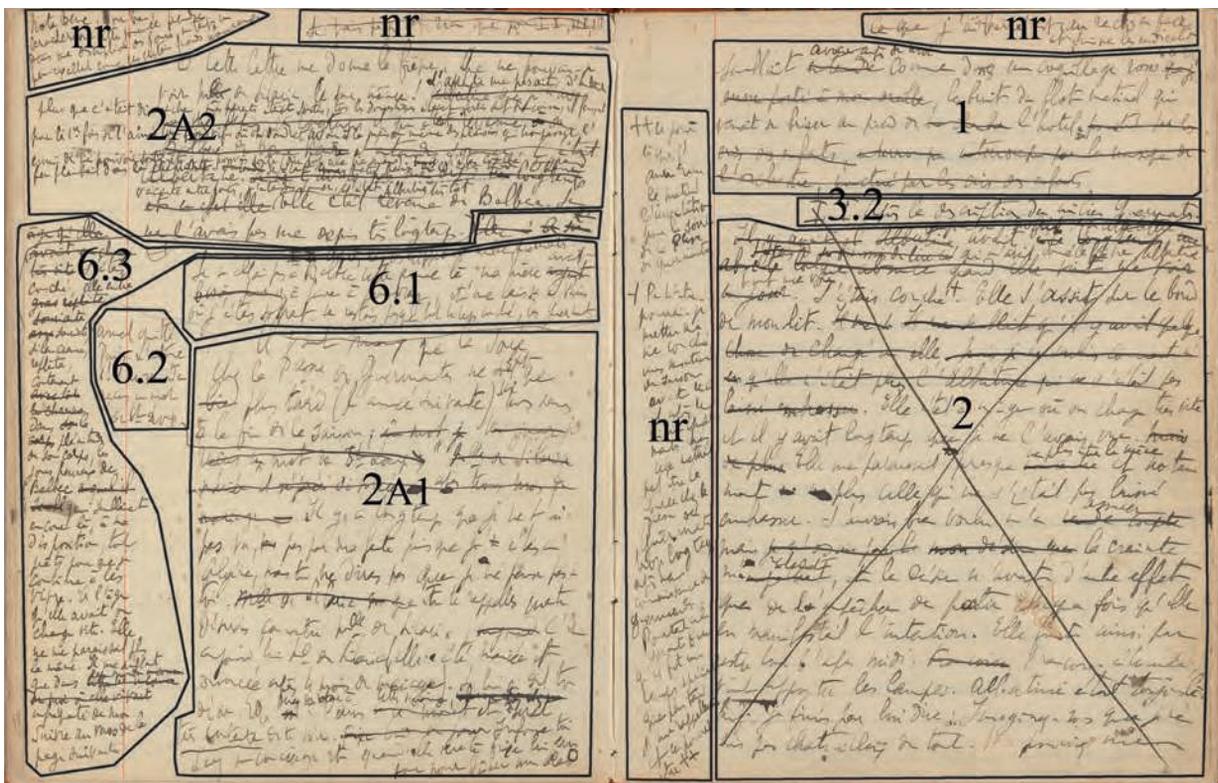
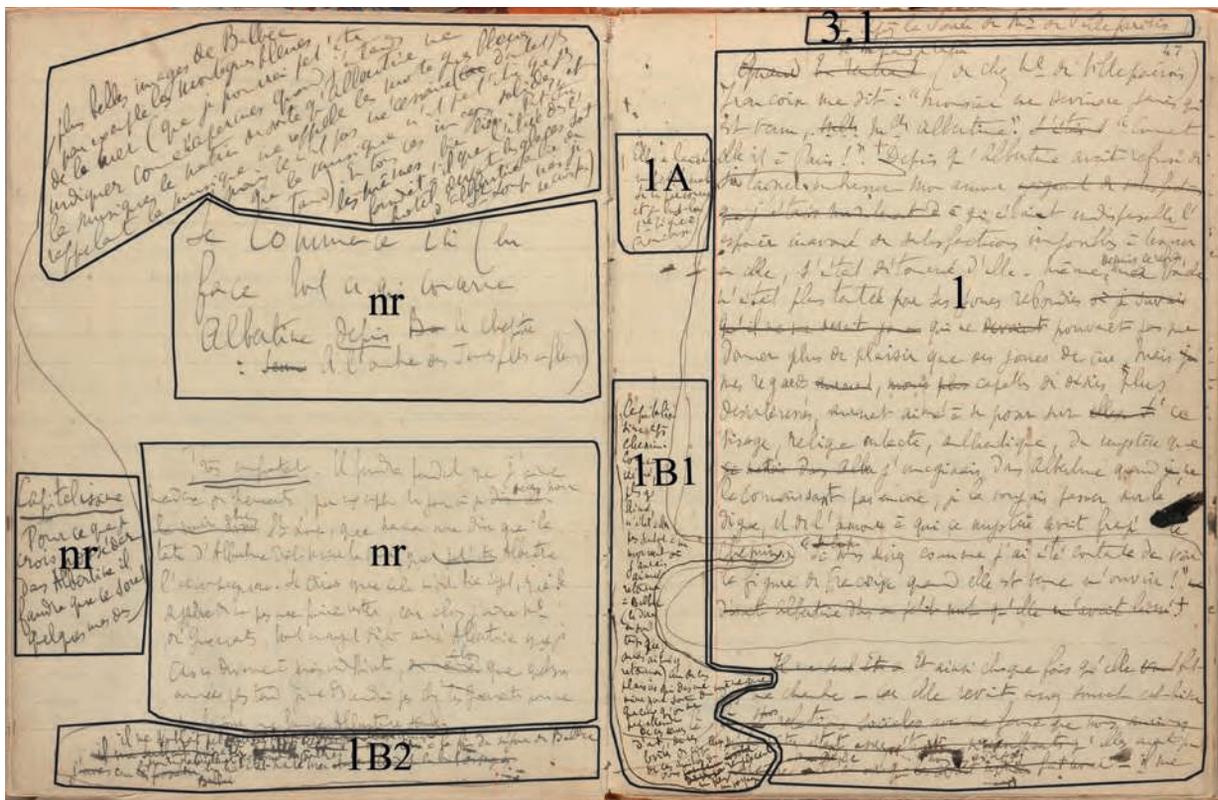


Fig. 1 : Cahier 46, f<sup>os</sup> 46 v<sup>o</sup>-50 r<sup>o</sup> ; séquence d'écriture (BnF, NAF 16686)



état. L'arrière-plan des zones a été colorisé en fonction du degré de certitude de l'éditeur concernant leur chronologie relative dans la séquence : plus la couleur d'une zone est foncée, plus le caractère conjectural de sa place dans la séquence est élevé. Grâce à ce codage visuel, l'utilisateur est informé de manière intuitive des doutes comme des convictions des chercheurs. Il peut également basculer à n'importe quel moment du mode « écriture » au mode « lecture ».

Le prototype est encore sommaire : par exemple, on ne peut pas représenter, pour le moment, les séquences dont les zones se répartissent sur plusieurs pages, comme celles que l'on voit sur la figure 1. Mais, selon nous, c'est un premier pas important vers la représentation du caractère dynamique et de la dimension temporelle qui restent trop souvent implicites dans les manuscrits de travail.

Les deux séquences proposées ici pour parcourir un fragment du Cahier 46 ne sont que des exemples. On pourrait imaginer bien d'autres types de parcours à travers les cahiers de Proust : thématique – bâtir une séquence sur toutes les visites d'Albertine au héros dans le Cahier 46,

par exemple – ou encore, chronologique – suivre la genèse d'un motif à travers différents cahiers. De fait, la notion de « parcours » semble particulièrement suggestive parce qu'elle propose une autre représentation de la genèse qui met l'accent précisément sur son aspect processuel. Et c'est sans doute cette dimension de l'écriture au brouillon que l'édition numérique pourra enfin prendre en charge de façon plus satisfaisante que l'édition papier, limitée par son espace bidimensionnel. L'encodage en TEI apparaît donc comme un outil fructueux, au sens où il rend possible à la fois la construction de ces parcours et leur visualisation.

---

17. Le prototype est disponible à l'adresse suivante : <[http://research.cch.kcl.ac.uk/proust\\_prototype/index.html](http://research.cch.kcl.ac.uk/proust_prototype/index.html)>. Il utilise une combinaison de *SVG* et *JQuery* générée à partir de XML via un script XSLT. *Scalable Vector Graphics (SVG)* : <[www.w3.org/TR/SVG/](http://www.w3.org/TR/SVG/)> ; *JQuery* : <<http://jquery.com/>> ; *eXtensible Stylesheet Language Transformation (XSLT)* : <[www.w3.org/TR/xslt20/](http://www.w3.org/TR/xslt20/)>.

---

**JULIE ANDRÉ** est Docteur en littérature et civilisation françaises de l'université Paris III et chargée de cours à Sciences Po Paris et à l'Université de Lorraine. Elle poursuit ses recherches littéraires et génétiques sur Proust au sein de l'ITEM (CNRS-ENS). Sa thèse, effectuée sous la direction de Pierre-Louis Rey, est consacrée à la transcription et à l'interprétation de l'un des cahiers de brouillon de Proust, le Cahier 46, dont elle prépare également l'édition dans le cadre de la collection des « Cahiers 1 à 75 de la BnF » (BnF/Brepols).

julie.andre@sciences-po.org

**ELENA PIERAZZO** est Docteur en philologie italienne de l'École normale supérieure de Pise et *Lecturer* en *Digital Humanities* au King's College de Londres, où elle dirige un Master en *Digital Humanities*. Elle occupe la fonction de *chair* de la TEI (*Text Encoding Initiative*) depuis 2012 et a été l'un des principaux contributeurs au développement d'un nouvel encodage pour les manuscrits modernes et les éditions génétiques. Elle a travaillé à plusieurs projets d'édition numérique, en particulier pour l'édition des manuscrits de Jane Austen et les premières éditions de Jonathan Swift.

elena.pierazzo@kcl.ac.uk

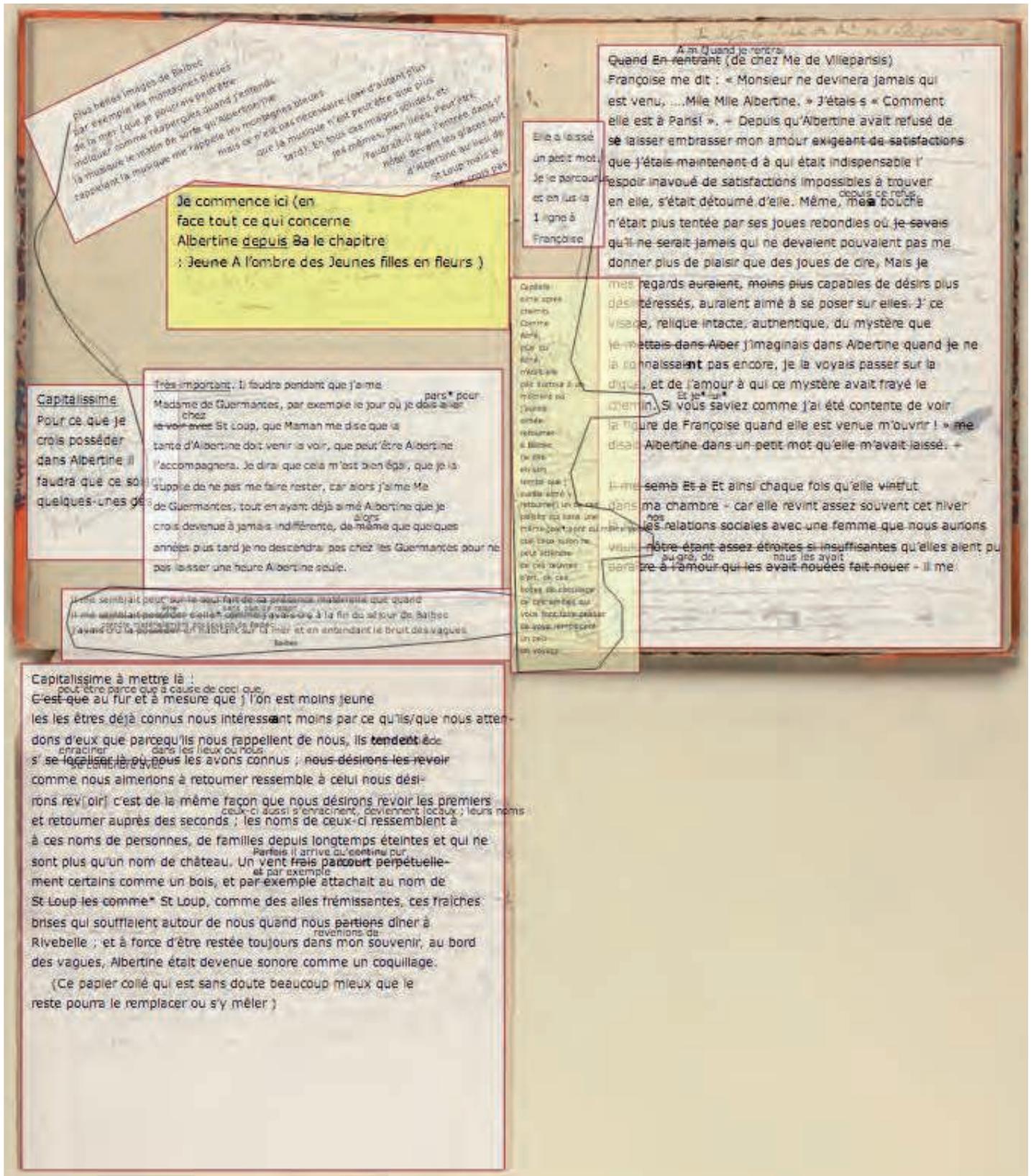


Fig. 3 : Capture d'écran du prototype pour la séquence de lecture dynamique des folios 46 v°-47 r° du Cahier 46, d'après <[http://research.cch.kcl.ac.uk/proust\\_prototype/](http://research.cch.kcl.ac.uk/proust_prototype/)>