

# QUEL BALISAGE POUR LES CORPUS ÉPISTOLAIRES NUMÉRIQUES ? De l'annotation traditionnelle du "document" à une analyse générique et pragmaticienne

Françoise LERICHE  
ITEM (CNRS), Université Grenoble 3

## SOMMAIRE

### 1. Enjeux, propositions

1.1. L'édition électronique et l'émergence de nouveaux corpus textuels et/ou documentaires

1.2. Le mode d'édition / d'annotation détermine le mode de lecture / d'usage

1.3. L'épistolaire : entre document historico-biographique et classe textuelle spécifique

1.4. De l'annotation traditionnelle (éclaircissements référentiels) à une annotation multi-critères, en vue de l'extraction de sous-corpus de recherche et d'exploitations statistiques

2. Rapide examen critique de corpus épistolaires numérisés ou numériques

3. Pour un balisage spécifique des corpus épistolaires modernes : illustration

3.1. Balisage formel ou balises de structure (constituants formels de la lettre)

3.2. Balisage thématique

3.3. Balisage des opérations pragmatiques

Bibliographie

Annexe 1 : lettres de Proust prises pour supports de l'analyse

## 1. Enjeux, propositions

Malgré les Cassandre du monde éditorial français, qui voient dans la concurrence de l'édition électronique la fin de l'édition (sélectivité, éthique, etc.), tirant argument de quelques entreprises vouées – en effet – à la diffusion hâtive et purement mercantile de textes mal édités (« édition sans éditeurs », lit-on çà et là), on peut de manière générale, en prenant un peu de recul, estimer qu'au contraire, les possibilités numériques sont en train d'introduire un nouvel essor et une mutation positive dans le domaine de l'édition et la philologie – comme l'imprimerie, à la Renaissance, avait permis l'apparition de nouveaux corpus et d'une nouvelle philologie (Rastier, 2001, p. 53 sqq.).

Parce que l'édition numérique permet de nouveaux modes d'accès aux textes (par rapport à la lecture linéaire du rouleau antique et même par rapport aux accès tabulaires du codex ou du livre imprimé – sommaires, index, etc.), la « philologie numérique » (Rastier, 1991, pp. 53-65), par diverses formes de balisages, facilite de manière inédite la lecture des textes et, mutation essentielle, révolutionne leur appréhension par la constitution de ces textes en « corpus » raisonnés. Mais le plus révolutionnaire (à mes yeux, du moins) réside dans le fait que l'ère numérique rend possible (est déjà en train de rendre possible) l'édition de textes qui résistaient à l'édition imprimée : fonds de manuscrits, vastes correspondances, œuvres multimédia, œuvres fictionnelles anciennes couvrant des milliers, voire des dizaines de milliers de pages... Ensembles textuels dont *seule* une édition électronique peut établir le texte, l'annotation, les modes d'accès et de recherche adéquats.

### 1.1. L'édition électronique et l'émergence de nouveaux corpus textuels et/ou documentaires

Les manuscrits<sup>1</sup> constituent un exemple typique de corpus laissés de côté par l'édition papier : fragmentaires par nature, raturés, offrant un ordre de lecture incertain voire incompréhensible, ils se refusent à la lecture linéaire qu'implique – malgré tout – le livre imprimé. Idéologiquement ils ont aussi, pendant des décennies, été négligés par la philologie classique, relégués à un vague statut de « documents », de traces documentaires, et non considérés comme des textes à part entière. Héritière de la philologie antique et médiévale, la philologie classique ne s'est intéressée aux « manuscrits » que pour l'établissement des textes, non pour étudier leur genèse, et les apparats

---

<sup>1</sup> J'entends par « manuscrits » non pas le manuscrit « au net », la copie d'impression, mais les manuscrits *de travail* d'un écrivain – autrement dit : les brouillons.

critiques des éditions « savantes » n'ont jamais rien fourni d'autre que les quelques « variantes » (lexicales ou grammaticales) des dernières étapes génétiques (manuscrit au net, copie d'impression, épreuves corrigées)<sup>1</sup>. Les manuscrits ne se laissant appréhender que sur un mode de lecture rhizomatique (et non linéaire), il fallait l'émergence d'un nouveau médium éditorial – l'édition hypertextuelle – pour rendre envisageable leur édition. La constitution de ces nouveaux corpus numériques est indissociable d'une réflexion approfondie sur les modes d'édition et d'annotation nécessaires pour rendre ces textes lisibles sans les dénaturer<sup>2</sup>.

Le cas qui nous retiendra ici, l'édition de fonds épistolaires, est également un de ces cas où les possibilités offertes par les outils numériques transforment radicalement l'approche du matériau : longtemps considérées comme fonds de documents historiques (traces biographiques, documents d'« accompagnement » de la genèse des œuvres), les correspondances peuvent désormais constituer des corpus textuels, si un mode éditorial spécifique parvient à désancrer les lettres de leur concaténation chronologique – donc de la logique historico-biographique des éditions papier qui, jusqu'à présent, a imposé ce mode de lecture essentiellement documentaire des correspondances.

### 1.2. Le mode d'édition / d'annotation détermine le mode de lecture / d'usage des textes

Ce principe, qui vaut en règle générale, s'avère particulièrement crucial pour les ensembles fragmentaires que sont les correspondances. Tandis que les romans, les recueils de poèmes, les pièces de théâtre, les autobiographies, sont des textes constitués par l'auteur lui-même en volumes recevant leur unité et leur légitimité de ce geste éditorial (destiné au public), les correspondances, qu'elles soient monographiques (Proust-Gallimard) ou générales (*Correspondance* de Marcel Proust), sont des artefacts éditoriaux. Outre les questions pratiques posées par la notion de « recueil » de lettres (voir section 2 ci-après), se pose un problème fondamental de « lisibilité » qui tient au type même de discours qui définit le texte épistolaire : les lettres, n'ayant pas été écrites en vue de leur publication, ne produisent pas leur propre système de référents (ne sont donc pas auto-suffisantes, comme peut l'être un poème) mais s'articulent allusivement à des conversations privées et à des référents communs à l'épistolier et à son destinataire, qui échappent à un lecteur extérieur à cet échange (et en particulier au lecteur moderne, qui ne baigne pas dans l'univers de références historiques, politiques, culturelles du moment). La « lisibilité » de la lettre dépend donc étroitement de son co-texte éditorial, qui conditionne la réception du lecteur :

-dans le cas (rare) d'une relation épistolaire prolongée entre deux correspondants, et conservée dans sa (quasi-)totalité, une publication monographique (ainsi, Proust-Gallimard) offre un « texte » suivi, un *texte dialogique* qui se présente à la réception du lecteur comme l'histoire d'une relation (amicale, conflictuelle, etc.) entre un écrivain et son éditeur. L'annotation servant à préciser quelques références, bien que philologiquement indispensable, est à peine nécessaire au lecteur, pris dans la dynamique de l'échange ;

-dans les cas beaucoup plus fréquents où le destinataire n'a pas conservé ses lettres et ne publie que celles de Proust, la « lisibilité » de cet échange tronqué (par exemple, les lettres de Proust à Lucien Daudet – voir Daudet, 1929) se heurte fréquemment à l'écueil de la référence. Quelques annotations fournies par le destinataire lors de la première publication de ces lettres permettent de reconstituer l'objet de la discussion, et/ou le contexte. Mais comment lire ces lettres isolées de l'échange qui leur donne pleinement sens ? Daudet ayant, en outre, censuré à la publication une grande partie des passages concernant les médisances mondaines, les confidences ayant trait à la vie privée, les reproches ou les effusions sentimentales, pour ne retenir que les discussions générales et surtout littéraires, le lecteur est amené à lire ces lettres (tronquées) comme un

<sup>1</sup> Même lorsque des éditions modernes entendent donner un aperçu des brouillons de l'œuvre, il s'agit d'extraits décontextualisés, isolés des passages hétérogènes qui les précèdent ou qui les suivent dans le brouillon, isolés également des étapes rédactionnelles qui les précèdent ou les suivent dans l'élaboration de l'épisode concerné, et surtout, d'extraits rendus lisibles (ratures supprimées, etc.). Voir, par exemple, les « esquisses » fournies à la fin de chaque volume d'*À la recherche du temps perdu*, édition dirigée par J.-Y. Tadié, Gallimard, « Bibliothèque de la Pléiade », 1987-1989.

<sup>2</sup> L'HyperNietzsche ([www.hypernietzsche.org/](http://www.hypernietzsche.org/)) constitue un exemple intéressant d'édition numérique de manuscrits fondée sur une philologie génétique adaptée à ce nouveau médium : à la fois archive virtuelle, corpus textuel (transcriptions établissant soigneusement le texte, avec ses ratures, ses additions interlinéaires, etc.), annotation savante, et guidage rhizomatique à travers les divers documents du corpus (chemins génétiques, chemins thématiques).

florilège de jugements de Proust (sur la création littéraire, sur son œuvre, sur un certain nombre de sujets d'intérêt général) – donc : comme un prolongement thématique ou métadiscursif du roman proustien ;

-cependant, dans la plupart des cas, les éditions scientifiques de correspondances se font sous forme de correspondances générales classées chronologiquement – et c'est aussi le cas de la correspondance de Proust : l'édition de monographies étant le fait des destinataires eux-mêmes dans les années qui suivirent la mort de l'écrivain, ces recueils à petits tirages sont devenus introuvables, et leurs lettres ont été reprises et intégrées dans la monumentale *Correspondance* en vingt et un volumes, classés chronologiquement<sup>1</sup>. Restituées à leur contexte historique et culturel par une annotation érudite qui identifie chacune des références littéraires, artistiques, politiques, mondaines, les lettres gagnent en signification et s'éclairent mutuellement (Proust tenant souvent des discours similaires à plusieurs correspondants auxquels il écrit le même jour). Mais ce mode d'édition, qui les classe chronologiquement, tend à faire des lettres des fragments d'un journal intime, des témoignages biographiques. Il induit une lecture documentaire, où la dimension éminemment relationnelle du texte épistolaire se perd.

### 1.3. L'épistolaire : entre document historico-biographique et classe textuelle spécifique

Pendant plusieurs décennies, peut-être parce que les écritures de l'intime étaient tenues en suspicion, les correspondances ont été tenues à l'écart du champ « littéraire », et tout juste considérées comme des documents à valeur essentiellement historique – pour l'établissement de biographies (genre lui-même déconsidéré). Depuis une ou deux décennies, par un prévisible mouvement de balancier, la revalorisation de l'intime a produit un intérêt spécifique pour l'épistolaire, considéré désormais comme genre littéraire (Kaufmann, 1990)<sup>2</sup>. Genre littéraire ? ou classe discursive particulière ?

Avant de se hâter de trancher la question, il convient de remarquer l'hétérogénéité du corpus, même dans une correspondance d'écrivain, et même d'un écrivain de la qualité de Proust. À partir de quel moment une lettre est-elle « littéraire » ? Une partie des lettres sont des billets relativement brefs et utilitaires, et la majorité des autres, même longues, discutent de questions pratiques, substituts de conversations téléphoniques<sup>3</sup> ; à une époque où il y avait plusieurs distributions de courrier par jour, et même des « petits bleus » (courriers par réseau pneumatique) qui arrivaient chez le destinataire en quelques minutes, une grande partie des lettres de Proust, « écrites au galop », ont la même fonction que nos courriers électroniques : demandes, réponses, poursuite d'une conversation orale, etc., elles ont une fonction relationnelle privée – et ne sont pas destinées, à l'inverse des lettres de la marquise de Sévigné ou de Bayle, à être lues dans les salons. En revanche, il est indéniable que ces objets textuels se caractérisent par une forme spécifique : en-tête et date (parfois), formule d'adresse toujours, formule d'adieu, signature, qui les classent comme « lettres » au premier regard, avant même la lecture du texte proprement dit. Inscription de la relation, et souvent des coordonnées du réel, ces indices formels, ainsi que la modalité généralement discursive, font de la lettre une parole *adressée*. Je propose donc de définir l'épistolaire simplement comme « classe de textes » spécifique (plutôt que d'un « genre littéraire »).

Afin que cette classe textuelle au fonctionnement particulier (texte écrit, mais doté d'un grand nombre de traits discursifs de l'oral) puisse être étudiée pour elle-même, il est nécessaire de constituer un corpus numérique. La vaste correspondance de Proust (plus de 6000 lettres, sans compter les lettres inédites réapparaissant occasionnellement) peut constituer un premier ensemble. La base gagnerait à être augmentée par d'autres correspondances de la même époque, permettant d'étudier régularités ou différences, et de s'interroger sur la notion de « style épistolaire ».

Il n'en reste pas moins que ces textes à forte composante discursive, sont fondamentalement inscrits dans un contexte historico-culturel qui est leur objet premier de référence, et que la richesse de l'expérience culturelle de Proust fait de sa correspondance une base documentaire

<sup>1</sup> Marcel Proust, *Correspondance*, texte établi, présenté et annoté par Philip Kolb, Plon, 1970-1993, 21 tomes.

<sup>2</sup> L'Association Interdisciplinaire de Recherche sur l'Épistolaire (A.I.R.E.) publie une revue semestrielle (*Bulletin de l'AIRE*, puis *Revue de l'AIRE*) depuis 1988.

<sup>3</sup> Proust a eu le téléphone très tardivement chez lui, l'utilisait peu, laissant ses domestiques répondre à sa place et communiquer ses messages, et il finit par s'en passer totalement pour ne pas être dérangé.

importante pour l'étude de la culture de la Belle-Époque et de la première guerre mondiale, ce qui constitue actuellement la principale motivation des demandes sociales pour la constitution d'une telle base numérique.

#### **1.4. De l'annotation traditionnelle (éclaircissements référentiels) à une annotation multi-critères, en vue de l'extraction de sous-corpus de recherche et d'exploitations statistiques**

La constitution d'une base « correspondance de Proust » se trouve ainsi au centre d'une double demande et d'un triple enjeu :

- d'abord, une forte demande sociale (et non pas uniquement des spécialistes de Proust) de disposer du texte numérisé de ces six mille lettres (ou 21 volumes) afin de faciliter les recherches en plein texte ;
  - corrélativement, une demande d'annotation précise des références, qui permette de saisir les allusions de l'épistolier ; donc, implicitement, d'une base documentaire
  - mais parce qu'une lettre n'est pas uniquement un objet documentaire historique, il serait judicieux de profiter de la constitution de cette base numérique pour proposer une annotation (un balisage) qui permette aussi, par la suite, de mener des recherches sur le « genre » épistolaire en tant que tel, et d'étudier, dans cette perspective, le « style » épistolaire proustien.
- Comment concilier ces demandes et ces enjeux ?

L'annotation de Philip Kolb étant extrêmement érudite et précise, il n'est pas difficile – et il est nécessaire – de l'introduire sous forme de notes et de liens hypertextuels, pour fournir au tiers-lecteur les éléments contextuels de référence. Mais permettre des recherches sur l'épistolaire proprement dit requiert de nouveaux types d'annotations. Lesquels ?

La question peut se reformuler ainsi : que peut-on vouloir chercher dans un corpus épistolaire ? comment l'utilisateur peut-il souhaiter interroger ce corpus ? Ces recherches étant encore inédites (la masse des 6000 lettres est ingérable sous sa forme papier actuelle, même pour les spécialistes...), il faut donc *imaginer* les requêtes possibles sur un corpus épistolaire numérique...

- Une partie des requêtes sera sans doute d'ordre thématique (retrouver aisément toutes les lettres traitant de politique, ou de musique, ou de littérature, ou du roman proustien, entre autres possibilités<sup>1</sup>..). Faut-il intégrer un logiciel de traitement de corpus, afin que par des requêtes lexicales (listes de mots) l'utilisateur repère des occurrences thématiques ? On examinera cette question ci-après (3.2.).
- Une étude thématique ne peut être productive que si elle est distributionnelle (avec quel correspondant tel sujet est-il plus fréquemment abordé ?)
- Une étude centrée sur le style épistolaire devrait pouvoir étudier les formules d'adresse, de politesse, d'adieu, les citations, et les comparer (analyse distributionnelle par correspondants, ou par périodes). Ce qui requiert un balisage de ces éléments de structure (voir 3.1.).
- Par ailleurs une lettre, parole « adressée », est toujours destinée à agir sur le destinataire, qu'il s'agisse de lui demander un service, d'influencer ses décisions, de s'excuser ou de se justifier d'une accusation, ou au contraire de l'amuser, de maintenir la relation par un échange de nouvelles, par la sollicitude : la dimension « expressive » de l'épistolier est subordonnée à une visée pragmatique. Quel genre d'épistolier est Proust ? On peut vouloir mener ces études statistiquement à travers l'ensemble du corpus, mais aussi de façon distributionnelle (étudier le type de relation entretenu avec tel ou tel correspondant). Seul un balisage des opérations pragmatiques permettrait ces études spécifiques au style épistolaire (voir 3.3.).
- Toutefois, le corpus épistolaire, je l'ai souligné plus haut, n'est pas homogène. Il est constitué de genres (ou sous-genres) conventionnels : lettre de condoléances, de félicitations, accusé de réception, ordre de ventes à la Bourse, lettre de vœux, dédicaces, lettres administratives, ces genres étant relativement codifiés – même si l'épistolier peut introduire des écarts par rapport aux normes sociales ; en revanche, d'autres genres épistolaires sont moins conventionnels, telles les nouvelles que l'on échange, les échanges de vues amicaux, les négociations avec l'éditeur. Les genres très codifiés imposant une attitude pragmatique particulière, l'utilisateur peut vouloir exclure du corpus certains types de lettres, pour ne pas générer de bruit ou ne pas fausser les résultats dans certaines requêtes ; ou au contraire, il peut souhaiter n'étudier que certains types très codifiés, pour voir si Proust y adopte le comportement socialement attendu...(et plus tard, si la

<sup>1</sup> Actuellement, l'index des noms de personnes ne permet que de retrouver des passages où il est question de tel ou tel homme politique, musicien, etc., ce qui est très limitatif.

base parvient à intégrer plusieurs correspondances, pour mener des études comparatives – rhétoriques, socio-linguistiques, etc. – sur certains genres épistolaires codifiés). Il faut donc un balisage multiple qui permette de sélectionner ou d'exclure dans l'ensemble du corpus des sous-corpus (par dates, par correspondants, par thèmes, par type de lettre), et qui permette d'interroger le corpus choisi en fonction d'un ou de plusieurs critères croisés.

## 2. Rapide examen critique de corpus épistolaires numérisés ou numériques

L'apparition de la TEI et du langage XML, qui décuplent les possibilités d'annotation et rendent pensable la production d'éditions web-centrées facilement accessibles, est très récente, et le caractère expérimental, dispersé, de leurs applications rend difficile une vue globale sur les corpus numériques en cours de constitution, notamment dans le domaine encore peu étudié de l'épistolaire.

L'examen critique de deux bases très différentes dans leurs objectifs comme dans leurs pratiques annotatives, Frantext et Arcane, me fournira quelques remarques méthodologiques sur les écueils et les limites de certaines pratiques éditoriales numériques.

L'objectif de Frantext étant, comme chacun sait, de constituer une base de données lexicales pour l'établissement du *Trésor de la Langue française*, les textes qui y sont intégrés, quel que soit leur genre, sont étiquetés uniquement selon une perspective lexicale et morpho-syntaxique (afin de permettre la recherche de fréquences lexicales ou de constructions syntaxiques, de voisinages et de co-occurrences). Un maniement adroit des listes de mots permet des repérages thématiques dans les textes (à condition que l'expression de l'auteur ne soit pas trop métaphorique ou allusive : voir 3.2.). Mais outre la difficulté de ces requêtes pour un usager non expert, et l'impossibilité fréquente de visualiser les occurrences dans des contextes élargis, la philosophie éditoriale de Frantext ne peut convenir à l'édition de lettres : comme le montre une recherche effectuée sur la correspondance de Flaubert, l'unité textuelle, pour Frantext, est l'item bibliographique, c'est-à-dire chaque tome de l'édition imprimée numérisée (ici, l'édition Conard des années 1920). À la différence des recueils de poèmes définis (selon un groupement esthétique ou chronologique) par l'auteur lui-même, les tomes d'une correspondance générale recueillie et éditée à titre posthume sont des artefacts éditoriaux : ils regroupent parfois quatre ou cinq années, parfois une année, de la correspondance de l'écrivain, pour des raisons aléatoires (il faut un certain nombre de lettres pour faire un volume éditorialement viable). On peut étudier dans Frantext si tel mot (« roman » par exemple) est plus souvent employé dans le tome 1 que dans le tome 2 ou dans le tome 12, mais quelle est la valeur de ces informations, dès lors que le découpage des lettres en tomes est aléatoire, et que les lettres de 1844 qui appartiennent au tome 1 pourraient tout aussi bien se trouver dans le tome 2 si Conard avait décidé de faire des volumes moins épais ? Outre que le séquençage du corpus par années réelles est impossible, une analyse distributionnelle par correspondants est également impossible. Numériser une édition papier existante et la considérer comme « un » texte nie la *spécificité de la lettre comme unité textuelle*. Même s'il est tentant de numériser au kilomètre une édition imprimée (c'est rapide, économique), il convient donc de refuser cette solution.

La correspondance de Bayle, en cours d'édition par Anthony McKenna<sup>1</sup>, propose au contraire un classement par lettre, chacune étant définie par son destinataire, son destinataire, sa date, son lieu d'expédition et de destination. L'utilisateur peut donc choisir de faire des recherches à travers l'ensemble du corpus, ou de sélectionner des sous-corpus (par épistolier, par destinataire, par date ou tranche chronologique, par lieu). Cette reconfiguration possible du corpus à chaque requête tient compte de la spécificité du fonds. Le balisage (par hyperliens de couleurs différentes) offre des annotations philologiques (ratures, additions, etc.) et des annotations référentielles (identification de personnes, d'œuvres). Un logiciel de cartographie permet de visualiser et de mesurer statistiquement les échanges épistolaires entre les correspondants, répartis en plusieurs points de l'Europe. L'objectif de cette base est documentaire et culturel : il s'agit d'étudier la diffusion de la pensée des Lumières à travers l'Europe du XVIII<sup>e</sup> siècle. Ces lettres, en général fort longues, soigneusement écrites, pleines d'informations savantes et de discussions scientifiques et philosophiques, sont en effet destinées majoritairement à être lues devant un public (parfois familial, souvent un cercle savant). Cependant, même si la relation intersubjective et les visées

<sup>1</sup> Édition numérique dans une base de données Arcane développée par Éric-Olivier Lochard à l'Université Jean Monnet de Saint-Étienne ; et, parallèlement, édition imprimée, chronologique et annotée, à la Oxford Foundation.

pragmatiques y tiennent moins de place que le contenu informatique (documentaire), ces traits spécifiquement épistolaires ne sauraient être absents de ces lettres – mais aucun balisage thématique ni formel n'est prévu, à ma connaissance, non plus qu'aucun outil permettant l'étude du lexique ou du style des épistoliers.

Comme pour l'édition imprimée (voir 1.2.), les présupposés scientifiques qui ont présidé à la constitution des bases numériques induisent la détermination de leurs outils de recherche et donc une certaine limitation dans l'utilisation de la base, au détriment d'autres questionnements auxquels ces textes pourraient, légitimement, être soumis.

### 3. Pour un balisage spécifique des corpus épistolaires modernes : illustration

Les réflexions qui suivent s'appuieront sur quelques exemples tirés de la correspondance de Proust (voir Annexe 1, exemples 1 à 3), pour interroger la pertinence des outils existants et proposer – de façon plus empirique que théorique – des modes de balisage spécifiques pour une étude non seulement historico-documentaire, mais aussi générique et culturelle de l'épistolaire.

#### 3.1. Balisage formel ou balises de structure (constituants formels de la lettre)

Dès lors que le balisage XML autorise une identification multiple d'un « document » – les balises « auteur », « destinataire », « lieu », « date », étant considérées comme balises de « structure » –, on peut étendre la nomenclature de ces balises de structure à l'étiquetage d'autres constituants génériques comme les formule d'adresse et d'adieu, la signature, les Post-scriptum. Un autre trait formel participe de la poly-énonciation épistolaire : les citations. Notre approche se veut ici purement descriptive : repérer les éléments formels entrant dans le texte de la lettre, qu'ils soient spécifiques de l'épistolaire ou également présents dans d'autres formes de texte. (Ainsi, le statut de la citation épistolaire gagnerait à être discuté d'un point de vue théorique : il y a plusieurs sortes de citations, et qui remplissent plusieurs fonctions ! Étudier si le régime citationnel des lettres en regard du régime citationnel romanesque requerrait un examen comparatif de ces deux types de corpus, ce qui supposerait des corpus numériques étiquetés de sorte à faire apparaître les citations. Au stade où nous en sommes, il est donc prématuré de décider si tel usage de la citation est littéraire et tel autre spécifiquement épistolaire. Toutefois, on peut remarquer que l'échange épistolaire introduit une dimension intertextuelle spécifique : la citation ou la reformulation des paroles du destinataire auxquelles l'épistolier répond – trait spécifiquement dialogique).

Exemple de balisage auteur/destinataire :

```
[MARCEL PROUST À ANTOINETTE FAURE] : [<name type=auteur> MARCEL PROUST</nom> À  
<nom type=destinataire> ANTOINETTE FAURE</nom>]
```

Exemple de balisage (simplifié) des formules d'adresse et de politesse :

```
Ma chère Antoinette : <adresse> ma chère <nom> Antoinette</nom>, </adresse>  
Mon cher petit grand-père : <adresse> Mon cher petit <nom>grand-père</nom>, </adresse>
```

Le balisage est ici simplifié pour les besoins de l'exposé. Il va de soi, notamment, que la balise « nom » doit préciser qu'Antoinette est Antoinette Faure ou que le grand-père est Nathé Weil, afin qu'une recherche des occurrences de « Nathé Weil » puisse sélectionner cette occurrence.

Exemple de citations qui doivent être balisées (afin de permettre des recherches sur l'univers intertextuel de Proust aussi bien que des recherches sur la polyphonie énonciative) :

```
soldat « simple et sublime » comme dit le Petit Boulangiste  
« Gais et contents nous allions triomphants »  
« C'est Boulange, lange, lange, »
```

Dans la lettre à Halévy (Annexe 1, exemple 3), la question se pose de savoir s'il faut baliser comme « citation » l'adjectif « décadent » souligné par l'auteur : « Je ne suis pas *décadent* » suppose que le destinataire a traité le jeune Proust de décadent, et qu'il s'agit là d'une reformulation allusive. Cet exemple montre la différence entre une description simple des typologies textuelles (texte citationnel vs texte de l'auteur) et une approche pragmatique des

phénomènes discursifs, qui ne se contente pas des marques formelles (guillemets) ou des signes explicites (« comme dit » X), mais qui interprète le texte épistolaire dans le contexte de la relation discursive des deux amis. Peut-être le recours à une balise « allusion » pourrait-il résoudre la différence entre ces deux types de pratique intertextuelle.

### **3.2. Balisage thématique**

Les requêtes thématiques constituent évidemment l'un des types de recherche privilégiés par les usagers. Les logiciels de traitement de corpus, s'ils sont très simples, ne permettent pas beaucoup de résultats et génèrent beaucoup de bruit ; s'ils sont complexes (comme Frantext), ils requièrent une compétence qui décourage même des étudiants avancés. Guider l'utilisateur dans le corpus nécessite un balisage thématique.

En tant qu'éditrice (putative) d'un corpus épistolaire, je souhaiterais pouvoir disposer d'outils de balisage sémantique automatisé, mais ces outils sont-ils fiables sur des textes qui, comme ceux de Proust, du fait de leur style allusif et métaphorique, nécessitent une large part d'interprétation ?

Les trois lettres de l'Annexe montrent des pratiques sémantiques hétérogènes : dans la lettre 2, les substantifs « somme » et « francs » (bien qu'ambigus pris chacun isolément) peuvent être, du fait de leur collocation, automatiquement étiquetés dans une nomenclature de type « argent » ; quant à « bordel », etc., il mène sans ambiguïté à un classement du passage dans « sexualité », par exemple. De même, dans la lettre 3, la redondance des termes « amour », « aimer », etc., « chair », « genoux », permet un classement automatique sous les rubriques d'« amour » (ou « sentiment ») et « sexualité ». – En revanche, au début de la lettre 3, seul le terme « poète » permet d'assigner mécaniquement l'étiquette « littérature » (ou « poésie ») à une discussion allusive qui multiplie les référents culturels. Et dans la lettre 1, aucun terme explicite ne permet de définir que la première séquence concerne la politique. En effet, « soldat » et « général » ont toute chance de produire un étiquetage dans la catégorie « guerre » (par exemple), de sorte que la lettre pourrait apparaître dans une requête concernant la première guerre mondiale, mais nullement dans une recherche de lettres relatives à la politique... Même un historien cherchant « Boulanger » ne pourrait tomber sur cette lettre, le nom de Boulanger ne figurant explicitement nulle part dans le texte. Ce n'est qu'un exemple parmi quantité d'autres, Proust étant maître dans l'art de l'allusion culturelle. Ne faut-il pas, dans ce cas, envisager un balisage manuel du corpus ? (C'est une question...)

### **3.3. Balisage des opérations pragmatiques**

La correspondance étant une pratique essentiellement interactive, relationnelle, le « style », la « rhétorique » du texte épistolaire ne peuvent s'appréhender, me semble-t-il, en dehors du cadre théorique de la linguistique pragmatique. L'épistolier est-il prompt à la récrimination, plaintif, toujours porté à demander des services, ou au contraire prompt à faire rire son correspondant ? ou a-t-il une tendance marquée à la confiance, à l'effusion lyrique, à l'expression de ses doutes et de ses craintes ? Seul un balisage spécifique permettra de mesurer objectivement, statistiquement, le « style » épistolaire, globalement ou de façon distributionnelle. Mais une fois énoncée cette évidence, comment opérer un balisage des opérations pragmatiques ?

Il faudrait pouvoir disposer d'une nomenclature stable, scientifiquement reconnue et acceptée, des « actes de langage ». Là réside un certain nombre de difficultés. Austin (1962) distingue les actes illocutoires par lesquels le locuteur « fait » quelque chose et les actes perlocutoires par lesquels le locuteur vise un certain effet. Tout d'abord, faut-il baliser les actes illocutoires, ou les actes perlocutoires (le but visé, la finalité) ? Mais la taxinomie des actes illocutoires est elle-même discutable. Ainsi, Austin divise ces actes illocutoires en cinq catégories : les verdictifs (exercice d'un jugement : décréter que, estimer, coter, juger, etc.) ; les exercitifs (exercice d'un pouvoir institutionnel de la parole : ordonner, excommunier, nommer, renvoyer, etc.) ; les promissifs (promettre, jurer de, donner sa parole, etc.) ; les comportatifs (catégorie assez disparate, de tout ce qui a trait au comportement social : remercier, s'excuser, compatir, critiquer, applaudir, etc.) ; et enfin les expositifs (catégorie « difficile à définir » de l'aveu même du théoricien, regroupant l'argumentation, l'explication : dire, affirmer, nier, concéder, etc.). Austin reconnaît cependant que sa classification est inachevée et imparfaite. – Si, pour la lettre 3 (Annexe) la catégorie des expositifs est satisfaisante pour la plupart des séquences (Proust expose sa façon de voir, argumente), dans le cas de la lettre 2, où ranger la « réclamation » ? La justification (il est impossible d'échouer deux fois) relève de l'argumentation, certes. Mais la « demande »

initialement formulée est un acte de langage spécifique, antérieur à toute argumentation, et doit pouvoir être catégorisé comme tel. En outre, que faire du micro-récit des déboires du jeune homme ? Selon Austin, un récit relève de la catégorie du constat, qu'il tient en dehors des actes illocutoires (« actes effectués *en* disant quelque chose, par opposition à l'acte *de* dire quelque chose » Austin, 1962, 113, Eluerd, 150). Pourtant, en racontant ses déboires, le jeune homme *fait* deux choses : il fait un aveu, il justifie sa demande d'argent : le récit n'est donc pas simplement un énoncé constatif. En outre, sa visée perlocutoire est claire : apitoyer son aïeul, pour obtenir la somme recherchée. En revanche, dans la lettre 1, le récit (les différentes séquences narratives) n'ont aucune valeur illocutoire particulière dans la classification d'Austin, mais cet égrenage de petits faits quotidiens « fait » quelque chose (donner des nouvelles) et a une fonction particulière : maintenir la relation entre amis pendant les vacances...

Une autre classification ne pourrait-elle pas mieux rendre compte de ces actes de langage ? Searle, critiquant Austin, propose quelques années après une taxinomie fondée non seulement sur le but illocutoire, mais aussi sur l'état d'esprit du locuteur et sur le rapport de son discours au monde. Cinq catégories également : les assertifs (affirmations, assertions, descriptions, caractérisation, explications, etc.) ; les directifs (essayer de faire faire quelque chose par l'interlocuteur : demande, ordonner, supplier, interdire, etc.) ; les promissifs (s'engager à faire quelque chose) ; les expressifs (exprimer un état psychologique : s'excuser, féliciter, remercier, etc.) ; les déclarations (provoquer un changement par une déclaration ; il s'agit de performatifs : démissionner, déclarer la guerre, excommunier, etc.). Cette taxinomie souple, qui peut paraître séduisante, a été très sévèrement critiquée par les théoriciens de la linguistique pragmatique, qui lui reprochent d'ignorer les usages ordinaires de la langue (voir Eluerd, p.165-175) : elle reproduit, en fait, les fonctions de la communication de Jakobson (fonctions référentielle, conative, expressive, etc.), est centrée sur le locuteur exclusivement, et ne tient compte ni du contexte, de la situation de communication, ni de l'interlocuteur.

Si dans une conversation orale les locuteurs sont, en effet, des co-locuteurs qui construisent ensemble la modalité de l'échange et son orientation, dans le cas des lettres, cependant, nous avons chaque fois affaire à une énonciation singulière (même si celle-ci est ouverte en permanence sur l'univers discursif des deux correspondants qui affleure en permanence dans le texte épistolaire).

Sans vouloir adopter la taxinomie de Searle, contestable, et en tenant compte de la visée implicite de tout « acte de langage » épistolaire, il me semble qu'il faudrait trouver, empiriquement, un étiquetage qui tienne compte à la fois de ce que « fait » la lettre et ce qu'elle « vise ».

Dans la lettre 1 (Annexe), Proust raconte : il donne des nouvelles (c'est ce qu'il fait), la visée étant (possiblement) de maintenir le contact, de partager une communauté d'opinion. Dans la lettre 2, en revanche, quand il raconte, il fait un aveu, et cet aveu a apparemment pour but de justifier sa demande (« voici pourquoi » : rhétorique de justification), mais en réalité d'apitoyer le grand-père. Un même « type » textuel, le récit, a dans ces deux *contextes* une fonction et une finalité différentes.

On pourrait donc imaginer une classification assez large de ce que « fait » l'épistolier (donne des nouvelles, confidence/aveu, exprime ses idées, polémique, donne une invitation/un rendez-vous, refuse une invitation, etc.), qui, couplée à une nomenclature thématique (scolarité, politique, amour, argent, littérature, etc.), permettrait des recherches relativement précises à travers le corpus.

Faut-il en rester là, et laisser à l'usager le soin d'approfondir l'analyse pragmatique à partir de sous-corpus qu'il aura sélectionnés ? C'est une option. Pourtant, il me semble qu'il serait intéressant de baliser aussi la *visée* des énoncés épistolaires. Ainsi, dans la lettre 2, la maxime (« il n'arrive pas deux fois dans la vie ».. etc.) a une finalité justificative ; de même, dans la lettre 3, les opinions littéraires émises par le jeune homme visent à le justifier d'accusations portées contre lui. Même chose pour la phrase finale, sur l'équivalence éthique de toutes les formes d'amour. Ce souci de justification me paraît récurrent dans la correspondance de Proust, et gagnerait à être quantifié.

## Conclusion

Ce programme d'édition et de balisage, quoique copieux, peut paraître insuffisant dans ses ambitions linguistiques, ne prévoyant pas d'inclure certains outils d'analyse lexicale, grammaticale, prosodique, etc. existant dans d'autres corpus numériques.



Reste à savoir si l'aide à la lecture doit viser l'exhaustivité. La base doit-elle permettre de tout faire ? Il me semble que dans un vaste corpus épistolaire, défini par sa double appartenance documentaire et textuelle, l'aide à la lecture consiste à fournir au lecteur des outils de repérage à travers l'hétérogénéité des lettres, afin qu'il puisse extraire les passages qui correspondent au type de recherches spécifiques à l'épistolaire qu'il est susceptible de vouloir mener dans ce corpus (recherches sur des genres épistolaires précis : lettres d'affaires, de condoléances, etc. ; recherches sur des catégories thématiques ou des pratiques culturelles : lettres sur la musique, citations, etc. ; recherches sur les interactions épistolaires : demandes de service, récriminations, attitude défensive, etc.) – ensuite, une fois extraits tous les passages pertinents, l'utilisateur ne pourra-t-il pas en étudier le style, le lexique, les caractéristiques énonciatives par ses propres moyens, ou les soumettre à d'autres outils d'analyse ?

Mais la discussion reste ouverte. Ce sont des propositions à caractère programmatique qui peuvent encore évoluer...

## BIBLIOGRAPHIE

- AUSTIN, J. L. 1962. *How to do things with words*, Cambridge, Mass, Harvard University Press, (*Quand dire c'est faire*. Introduction et traduction de G. Lane. Paris, Le Seuil, 1970).
- DAUDET, L. 1929. Autour de soixante lettres de Marcel Proust, *Les Cahiers Marcel Proust*, n° 5, Paris, Gallimard.
- ÉLUERD, R. 1985. *La Pragmatique linguistique*, Paris, Nathan.
- KAUFFMANN, V. 1990. *L'Équivoque épistolaire*, Paris, Minuit.
- KERBRAT-ORECCHIONI, C. 1990, 1992, 1994. *Les Interactions verbales*, Paris, Armand Colin, 3 tomes.
- PROUST, M. *À la recherche du temps perdu*, Édition publiée sous la direction de Jean-Yves Tadié, Paris, Gallimard, 1987-1989, Bibliothèque de la Pléiade, 4 tomes.
- PROUST, M. *Correspondance*, Texte établi, présenté et annoté par Philip Kolb, Paris, Plon, 1970-1993, 21 tomes.
- PROUST, M. GALLIMARD, G. *Correspondance 1912-1922*, Édition établie, présentée et annotée par Pascal Fouché, Paris, Gallimard, 1989.
- RASTIER, F. 2001. *Arts et sciences du texte*, Paris, Presses universitaires de France.
- RASTIER, F. 1991. *Sémantique et recherches cognitives*, Paris, Presses universitaires de France.
- SEARL, J. R. 1982. *Sens et expression. Études de théorie des actes de langage*, Traduction et préface de J. Proust, Paris, Minuit, (*Expression and Meaning*. New-York, 1979).

## ANNEXE

### Exemple 1 :

[MARCEL PROUST À ANTOINETTE FAURE]

[Paris le 15 juillet 1888]

Ma chère Antoinette,

Croiriez-vous que Maman m'a déchiré une lettre pour vous. L'écriture était trop mauvaise. Au fond je crois qu'un grand éloge de notre brave général, du soldat « simple et sublime » comme dit le Petit Boulangiste a excité les vieux sentiments orléanistes-républicains de madame Jeanne Proust. **[N1]**

Jamais les rues d'Auteuil (où j'ai passé seulement la journée du 14) n'avaient été aussi animées qu'hier. Vous ne trouvez pas entraînant ce refrain :

« Gais et contents nous allions triomphants »

ou :

« C'est Boulangé, langede, langede, »

hurlé par tous, femmes, ouvriers, jusqu'aux petits enfants de cinq à huit ans qui le chantent très très juste – avec ardeur.

Quoique l'homme soit très commun et un vulgaire batteur de grosse caisse, ce grand enthousiasme si imprévu, si *roman* dans la vie banale et toujours la même, remue dans le cœur tout ce qu'il y a de primitif, d'indompté, de belliqueux. **[N2]**

Vous voyez que je ne suis pas grand philosophe et je ne trouve guère que des adjectifs quand je cherche des raisons qui (pardonnez-moi cette enfilade de qui) me donnent envie de brailler : Il reviendra. **[Métadiscours]**

Je n'ai rien à vous raconter des Champs-Élysées. Blanche est toujours très douce, d'un visage angélique espiègle et résigné. Marie Bénardaky est très jolie et de plus en plus exubérante. Elle s'est *battue* à coups de poing avec Blanche qui a été *battue* et qui (ceci n'a pas de rapport) vous fait beaucoup remercier de votre lettre. **[N3]**

J'ai composé avant-hier au concours cinq heures de suite sans l'ombre d'un repos. Je suis arrivé à la Sorbonne à 9 h. 30 et j'en ai quitté à 4 heures moins un quart. La composition a duré de 10 h. 30 à 3 h. 30. Nous étions 120 ou 130 composants, c'est-à-dire les deux (rarement les trois) premiers de toutes les divisions de tous les lycées. C'était en histoire. **[N4]** J'irai entendre Paulus un de ces soirs. Je vous rendrai compte de la représentation. **[N5]**

Faut-il dire quelque chose de votre part à vos amies ? Je vais à peu près tous les jours aux Champs-Élysées. **[Offre de services, motivée]**

Je vous souhaite de bonnes et charmantes vacances ainsi qu'à toutes celles de vos amies que je connais.

Présentez mes affectueux respects à Monsieur et Madame Faure, à Mademoiselle Lucie, Mademoiselle Marcelle et sa sœur etc.

Marcel Proust.

**N1-N2** : nouvelles personnelles [N1] et collectives [N2]. **Thèmes** : la politique, le boulangisme

**N3** : nouvelles collectives (cercle restreint). **Thème** : anecdotes sur des amies communes

**N4** : nouvelles personnelles. **Thème** : le concours général, la scolarité de Proust

**N5** : nouvelles personnelles (projet). **Thème** : le café-concert, les arts du spectacle

Ces séquences sont coupées par des remarques de l'épistolier sur son style **[métadiscours]** et se terminent par une **[offre de service]**, suivie des formules d'adieu d'usage dans un texte de type épistolaire.

**Exemple 2 :**

[MARCEL PROUST À NATHÉ WEIL]

Jeudi soir [17 mai 1888]

Mon cher petit grand'père,

Je viens réclamer de ta gentillesse la somme de 13 francs que je voulais demander à Monsieur Nathan, mais que Maman préfère que je te demande. **[D]** Voici pourquoi. J'avais si besoin de voir une femme pour cesser mes mauvaises habitudes de masturbation que papa m'a donné dix francs pour aller au bordel. Mais 1° dans mon émotion j'ai cassé un vase de nuit, 3 francs 2° dans cette même émotion je n'ai pas pu baiser. Me voilà donc comme devant attendant à chaque heure davantage 10 francs pour me vider et en plus ces 3 francs de vase. Mais je n'ose pas redemander sitôt de l'argent à papa **[récit, à valeur explicative]** et j'ai espéré que tu voudrais bien venir à mon secours dans cette circonstance **[D']** qui tu le sais est non seulement exceptionnelle mais encore *unique* : il n'arrive pas deux fois dans la vie d'être trop troublé pour pouvoir baiser [...] **[maxime à finalité justificative]**

Je t'embrasse mille fois et n'ose te remercier d'avance. **[R]**

Je passerai demain à onze heures chez toi. **[rendez-vous]** Si ma situation t'a ému et que tu te rendes à mes prières j'espère que je te trouverai ou un commissionnaire chargé de la somme. **[D'']** En tous cas merci car ta décision n'aura pour cause que ton amitié pour moi [...] **[R']**

Marcel

**D = demande**

**D' = réitération de la demande**

**D'' = seconde réitération de la demande**

**Récit (ou confidence) : différence pragmatique avec N1, N2 etc. de la lettre 1. Ici, la narration a une valeur explicative, sert à motiver la demande D (On pourrait aussi bien intituler cette section : explication, ou motivation).**

**R : remerciement (on passe ici sur la litote)**

**R' : réitération du remerciement.**

**Exemple 3 :**

[MARCEL PROUST À DANIEL HALÉVY]

[Avant le mardi 22 mai 1888]

Mon cher Daniel

On géographise avec zèle autour de moi. Je me donne deux minutes de répit –. Je ne suis pas *décadent*. Dans ce siècle j'aime surtout Musset, le père Hugo, Michelet, Renan, Sully Prud'homme [sic], Leconte de Lisle, Halévy, Taine, Becque, France. Je me plais beaucoup à Banville, à Hèrèdia [sic] et à une certaine anthologie <idéale> composée de morceaux exquis de **poètes** que je n'adopte pas en entier : La Création des Fleurs de Mallarmé, des Chansons de Paul Verlaine etc. etc. – Mais j'ai horreur des critiques qui ont une attitude ironique vis-à-vis des décadents. Je crois qu'il entre dans leur cas beaucoup d'insincérité, mais inconsciente ou au moins sans clairvoyance. Les causes de cette insincérité sont si tu veux, la religion des belles formes de langage, une perversion des sens, une sensibilité malade qui trouve des jouissances très rares dans de lointaines accordances, dans des musiques plutôt suggérées que réellement existantes.

Quant au **style** Mendès, Silvestre, Banville (en **prose**) je crois qu'il mène à l'insincérité qui est le commencement de la banalité (ça ressemble un peu à M. Purgon : la dissenterie etc. etc.) : Si ça ne te semble pas très clair je t'expliquerai de vive voix. –

Je n'ai pas de **passion**. Je trouve ton ami le plus cher ou celui qui t'aime plus que les autres, je ne sais pas comment cela se dit en français – très gentil et j'ai un très réel plaisir – que je n'essaye pas de dissimuler – à me trouver avec lui. Mais comme j'en ai autant à me trouver avec toi, je voudrais que tu me dises quels jours tu ne rentres pas immédiatement à 4 heures. Et ceux où tu voudras bien de moi, je serai à tes ordres. Bizet trouvera que je commence avec toi la série des « listes » qui sont comme tu sais, chez moi le commencement de l'**amitié**. Il se tromperait absolument, car la mienne pour toi est déjà ancienne et quoique ce soit assez bête à dire mais enfin sur le papier tout passe, très vive.

Je te remercie de m'avoir donné cette occasion de ne pas écouter Choublier. D'ailleurs en t'écrivant je croyais te causer et je me suis ainsi donné l'illusion d'un grand plaisir, ce qui tu sais, n'est pas plus illusoire.

Post Scriptum

Je te propose de fonder avec moi (mais soyons seuls, directeurs) un grand **journal d'art**. –. Quant à ton **pédéraste** virtuel ou non, tu peux très bien te tromper. Je sais... qu'il y a des jeunes gens... (et si ça t'intéresse et que tu me promettes un *secret absolu*, même pour Bizet, je te donnerai des pièces d'un intérêt très grand à ce point de vue, à moi appartenant, à moi adressées) des jeunes gens et surtout des types de huit à dix-sept ans qui **aiment** d'autres types, veulent toujours les voir (comme moi, Bizet) pleurent et **souffrent** loin d'eux, et ne désirent qu'une chose les **embrasser** et se mettre sur leurs **genoux**, qui les **aiment** pour leur **chair**, qui les couvrent des yeux, qui les appellent **chéri**, mon ange, très sérieusement, qui leur écrivent des lettres **passionnées** et qui pour rien au monde ne feraient de la **pédérastie**.

Pourtant généralement l'**amour** l'emporte et ils se **masturbent** ensemble. Mais ne te moque pas d'eux et de celui dont tu me parles, s'il est ainsi. Ce sont en somme des **amoureux**. Et je ne vois pas pourquoi leur **amour** est plus malpropre que l'**amour** habituel.