

# De cahIEr à cahieR

## Pour des études littéraires assistées par ordinateur

---

Ioana Galleron, Fatiha Idmhand, Cécile Meynard

**PRESENTATION DU 8 JUILLET 2016**  
**SÉMINAIRE ANNUEL CAHIER, CAEN.**

# Etat des lieux

---

# La recherche en lettres à l'ère du numérique

---

- Editions en ligne = mise à disposition de textes rares, de collections complètes, de manuscrits
- Organisation des informations pour faire émerger de nouveaux objets de recherche

## MAIS

*‘Les ELAO\* [n] ont souvent comme résultat [que] de signaler des pistes de recherches, plutôt que d’effectuer l’ensemble d’une étude’. (M. Bernard, 2003)*

\*ELAO = études littéraires assistées par ordinateur

# La recherche en lettres à l'ère du numérique

---

Que nous a apporté l'aventure électronique jusqu'à ce jour?

- Dans quelle mesure voyons-nous nos corpus différemment?
- Que peut-on faire maintenant?
- Sommes-nous parvenus à répondre à des questions auxquelles nous ne pouvions pas répondre auparavant?
- Quelles nouvelles questions émergent?
- Comment modifient-elles les pratiques (de recherche, d'édition, d'informatisation)?
- **Sommes-nous assez outillés** pour y répondre?

# La recherche en lettres à l'ère du numérique

---

Actuellement : prédominance des projets qui donnent la priorité à la production/conversion de données “textuelles” en données numériques >>>>

- insistance sur le développement de certains langages de description (XML, XML-TEI, XML-EAD, Dublin Core, RDF, etc.)
- des outils *user (un)friendly* du côté de l'exploitation, surtout basés sur des usages « TAL » et sur des approches lexico/textométriques, archivistiques ou muséales.

# La recherche en lettres à l'ère du numérique

---

## Comment aller plus loin pour:

- exploiter les informations numérisées?
- croiser les corpus, les approches? confronter les données générées dans le cadre de projets différents?
- créer par nos travaux de la véritable valeur ajoutée?

Comment aller vers de  
véritables *ELAO*?

---

# Retour sur les résultats des projets Boissy, CHispa, Stendhal, Benoîte Groult

---

Projets portant sur des corpus littéraires allant du XVIII<sup>e</sup> au XXI<sup>e</sup> siècle

1. Louis de Boissy (1694-1758)
2. Stendhal (1783-1842)
3. José Mora Guarnido (1894-1967) & Carlos Denis Molina (1916-1983)
4. Benoîte Groult (1920-2016)

# Retour sur les résultats des projets Boissy, CHispa, Stendhal, Benoîte Groult

---

Pour ces seuls projets, publication de :

- + 500 œuvres (éditées ou inédites) de théâtre, fiction, poésies, etc.
- + de 1000 numéros de journal et articles de presse
- + de 25000 images en ligne
- + de 5000 pages de manuscrits avec leurs transcriptions

**Formats:**

- **Bases de données, avec des interfaces outil ou créées ad-hoc;**
- **éditions XML/XML-TEI + métadonnéesinteropérables, etc.**

# J'édite, donc je pense

---

La perspective initiale voulait transférer les repères de l'édition papier « classique »: approches philologiques, génétiques, critiques, etc.

Les projets visaient le :

- codage des variantes, des processus d'écritures,
- codage des particularités typographiques (ex. s long)
- maintien de l'orthographe d'origine ET normalisation (double accès proposé au lecteur);
- l'ajout d'autant de notes que souhaité
- Restructuration dynamique, différentes modalités d'accès et de visualisation des données

**mais .....**

# J'édite, donc je pense

---

le transfert vers le numérique a généré de nouvelles questions sur :

➤ les index: compte tenu de la démultiplication des possibilités?

➤ les attributs ou balises?

<name> + @type ou <persName>

<place>, <title>: quelle normalisation pour @n?

Xml:id? A quel niveau?

➤ le repérage dans une édition électronique?

✓ numéroté les vers, bien entendu...mais que fait-on des pages? Comment cite-t-on?

# J'édite, donc je pense

---

➤ la visualisation des résultats... mais comment?

Ex.

- circulation de la parole dans une pièce de théâtre
  - emboîtements d'espaces référentiels dans les œuvres de fiction
  - graphiques et structures métriques dans un poème
  - Organisation chronologique, thématique, etc. de documents  
etc.
- les standards d'une « bonne » édition électronique (de la même façon que l'on parle d'une « bonne » édition papier) ?

# J'édite, donc je pense

---

- Quel niveau de codage attendre en vue d'une réutilisation de l'édition par d'autres?
  - vision minimaliste: texte mis à disposition avec quelques métadonnées et mots clés
  - vision maximaliste: texte lourdement annoté (entités nommées, variantes orthographiques, variantes auctoriales, informations sur la nature grammaticale des mots, ponctuation, etc.) ; avec nombreuses métadonnées...
- Comment générer de nouveaux standards, et sur quels critères?

# J'édite, donc je transforme et je réinvente

---

C'est une nouveauté du numérique !

1. Les textes deviennent des données
2. Les informations extraites génèrent de nouvelles lectures.

Questions :

- Peut-on associer micro-lectures et macro-analyses assistées par ordinateur?
- Comment interpréter ces nouvelles RECIPROCITES (nouvelles lectures et interprétations, nouvelles circulations des textes, nouveaux croisements)

# Splendeurs et misères de la TEI

---

Les recommandations « minimales » de la TEI

```
<castList>
```

```
  <castItem>
```

```
    <role xml:id="neptune">NEPTUNE.</role>
```

```
    <actor>LE ROI.</actor>
```

```
  </castItem>
```

```
</castList>
```

# Splendeurs et misères de la TEI

---

Pourtant, on peut en savoir beaucoup plus!

```
<castItem ana="#OCH #SM #AGY #ARL">  
  <role>MONTVAL</role>,  
  <roleDesc>  
    <fs><f name="occupation">Officier</f></fs>,  
    cru <fs><f name="occupation">Médecin</f></fs>  
    fous le nom de M. <name type="person"  
      n="Bromps">Bromps</name>.  
  </roleDesc>  
</castItem>
```

# Splendeurs et misères de la TEI/de l'encodage XML

---

Et avec d'autres types de travaux, on en saurait plus sur :

- Les types de scènes
- Les topoi
- Les figures de style
- Les faits de langue(s) : *sabir...*

**Conclusion : la TEI ne permet pas tout et, surtout, de nombreuses tâches pourraient être systématisées !**

 Reste la question de l'extraction de l'information??

# Des outils pour interroger, oui, mais....

---

De quels outils disposons-nous vraiment?

➤ TXM, PhiloLogic, AntConc, etc. bref, des concordanciers en folie!

**MAIS les études littéraires ne travaillent pas qu'avec les mots!**

Les objets d'étude de la littérature requièrent des approches sémasiologiques !

➤ Comment identifier des macro-unités de sens?

➤ Comment intégrer les « lectures des lectures », donc la critique!

# Des outils pour exploiter

---

Ils sont mal connus et complexes à prendre en main ! (BaseX, Python, suite R...)

## **Conséquence :**

Le littérateur doit compter en permanence sur quelqu'un pour explorer son/un texte et les données obtenues.

## **Comment avancer?**

La formation aux outils, c'est bien, mais le partage, l'échange autour de tâtonnements, c'est plus que nécessaire !

Quelles interfaces dédiées?

# Coordonner les actions

---

Si en Humanités Numériques, une hirondelle ne fait pas le printemps....

Si une comédie (un auteur) encodé(e) ne suffit pas....

Alors, il faut :

ALIGNER/COMPARER/CONFRONTER

Big data = Big money ....

OU ...

consortia de projets (similaires)

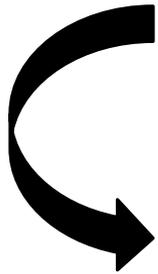
# Propositions

---

# Revoir ses copies...

---

1. Le numérique affecte la construction du savoir: l'informatique n'est donc pas qu'utilitaire !

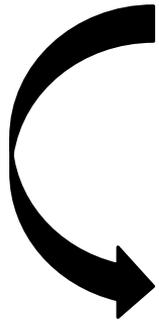


Il faut cesser la compilation d'outils pour aller vers « le dialogue d'égal à égal auquel incitent les ELAO, entre deux disciplines qui ont des choses à apprendre l'une de l'autre. »  
(Michel Bernard, 2003)

# Revoir ses copies...

---

2. Le chercheur a intérêt à aller sur le terrain de l'informatique :

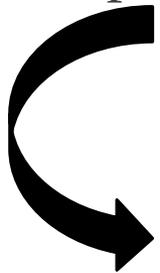


- pour tirer parti de la puissance des outils informatiques!
- pour s'adapter à une interface et à une base de données particulière
- pour réaliser davantage d'expériences et élaborer de nouvelles pistes théoriques

# Revoir ses copies...

---

3. L'utilité d'une base de données ne réside pas dans la masse, mais surtout dans la nature et la qualité des informations qu'il est possible de croiser



Les méthodes doivent penser le traitement des informations et adapter les réponses (exploiter des métadonnées, comprendre ce qui fait sens d'un point de vue statistique, etc.)

# Revoir ses copies...

---

4. La langue littéraire est complexe (poésie, figures de style, etc. ?)



- Où est la méthode pour des études thématiques ou sémantiques assistées (#quantitatifs de la lexicométrie) par ordinateur?

# Donc...

---

5. La formation des chercheurs doit aller plus loin que l'enseignement des outils !

# Propositions pour une nouvelle épistémologie...

---

- Premier point : le mode de constitution et la taille du corpus ont changé d'échelle
- Deuxième point : les sources critiques sont accessibles et internationales (elles aussi ont changé d'échelle)
- Troisième point : les études littéraires doivent introduire l'expérimentation dans leur approche

# Une expérience de lecture... pour commencer...

---

Autour du personnage. Remettre le « personnage en questions » permet de traiter différents :

- ✓ genres : théâtre, roman, nouvelle, poésie, etc.
- ✓ courants
- ✓ périodes
- ✓ pays

➤ Mais également de :

- ✓ renouveler des approches théoriques à partir :
  - d'une « recherche-action » associant étudiants et chercheurs
  - d'une « recherche appliquée » : démarche expérimentale

# Le personnage remis en questions

---

## ➤ Objectif :

- ✓ Asseoir les bases d'une nouvelle conception du « personnage »
- ✓ Exploiter des données informatiques issues de résultats de lectures
- ✓ Construire de nouvelles théories littéraires:
  - en enrichissant les outils informatiques des approches sémasiologiques
  - en enrichissant les théories littéraires de résultats informatiques

# Le personnage remis en questions

---

## ➤ Comment?

En élargissant le groupe de travail « R » à :

- ✓ Membres de Cahier volontaires
- ✓ Extérieurs à Cahier (UDPN, Paris 3?, etc.) + « Liste DH »
- ✓ Wiki sur la liste Cahier (Renater)

## ➤ Calendrier ?

- ✓ Réunions tous les trois mois
- ✓ Bilan de l'expérience et des résultats à l'occasion de l'Atelier annuel de CAHIER 2017

# Le personnage remis en questions

---

➤ Quelles méthodes?

✓ Combiner une lecture de mille textes avec mille lectures d'un texte !

➤ Comment?

✓ Expérience autour d'une « population spécifique » : des étudiants issus des filières *études littéraires* de licence et master de quatre à six universités différentes

✓ Mise en place de questionnaires, wiki, listes d'outils pour faciliter la lecture, la prise de notes, la PAO, l'encodage, etc..

✓ Mise en place d'un espace de travail

✓ Propositions de développements informatiques

✓ Publication et exploitation des résultats des lectures

# Contacts

---

Ioana Galleron: [ioana.galleron@gmail.com](mailto:ioana.galleron@gmail.com)

Fatiha Idmhand: [fatihaidmhand@yahoo.es](mailto:fatihaidmhand@yahoo.es)

Cécile Meynard: [cecile.meynard@gmail.com](mailto:cecile.meynard@gmail.com)

Et nous suivre : <http://cahier.hypotheses.org/>

# Bibliographie...

---

A consulter sur zotero :

[https://www.zotero.org/groups/r\\_research\\_group\\_of\\_cahier](https://www.zotero.org/groups/r_research_group_of_cahier)