

Personnages de pièces de théâtre: codage et analyse

Séminaire Hubert de Phalèse

24 octobre 2017

Quelques mots d'introduction

- Spécialisation en littérature française des XVII^e et XVIII^e siècles (théâtre et poésie mondaine des années 1680 à 1740)
- « Mes » humanités numériques:
 - Encodage en XML/TEI
 - Extraction de données avec BaseX (Xqueries)
 - Un peu de textométrie (AntConc, TXM, Hyperbase)

Quelques mots d'introduction

- Le personnage: un cadre théorique convaincant, mais tremplin de discours fondés surtout sur l'intuition et la sélection des arguments
- Pas ou peu de connexion entre les cadres théoriques (structuralistes ou sémiotiques) et les études menées sur les personnages (du moins en ce qui concerne le théâtre des siècles classiques)

Cadres théoriques: Hamon

- Personnage défini par contraste avec d'autres personnages
 - qualification différentielle (prédicats)
 - distribution différentielle (apparitions)
 - autonomie différentielle (seul ou accompagné)
 - fonctionnalité différentielle
 - pré-désignation conventionnelle
 - commentaire explicite (portrait)

Cadres théoriques: Jouve

- « effet-personnage »
 - Effet personnel
 - Effet personne
 - Effet prétexte

Quelle possibilité d'opérationnalisation des deux premiers dans les textes littéraires en général, et les pièces de théâtre en particulier?

Quelles nouvelles analyses?

Plan de l'intervention

Retours d'expérience sur:

- 1° l'analyse des personnages par analyse avec des logiciels de textométrie
- 2° l'encodage des personnages dans les éditions XML/TEI

Personnage et textométrie

- Une pléthore d'outils, mais peu d'emplois pour l'analyse littéraire: attribution auctoriale, analyse stylistique
- Travaillons-nous sur les mots?
- Analyse du système très codifié de la comédie classique (1630-1810): principe du kaléïdoscope
- Description et catégorisation des éléments kaléïdoscopiques

Personnage et textométrie

- Peut-on classer les personnages selon les caractéristiques de leurs discours?
- Intérêt moins pour la polarisation lexicale (niveaux de langue), mais pour la distribution des unités les plus fréquentes
- « Indice pronominal » de Charles Muller > travail sur les « pronoms du dialogue » (je, tu, vous)

Personnage et textométrie

8 pièces encodées en XML/TEI

- *La Galerie du Palais* (1632), de Pierre Corneille ;
- *L'École des femmes* (1662), de Molière ;
- *Le Chevalier à la mode* (1687) de Saint-Yon et Dancourt ;
- *Le Joueur* (1696), de Regnard ;
- *Ésope à la cour* (1701) de Boursault ;
- *La Surprise de l'amour* (1722), de Marivaux ;
- *Le Glorieux* (1732) de Destouches ;
- *Les Dehors trompeurs* (1740) de Boissy

Personnage et textométrie

Discours de chaque personnage (<castItem>) isolé avec une requête Xquery

```
//*:sp[@who='nom_personnage']/*:p/text()  
| //*:sp[@who='nom_personnage']/*:p/*:s/text()  
| //*:sp[@who='nom_personnage']/*:p/*:name/  
text()  
| //*:sp[@who='nom_personnage']/text()  
| //*:sp[@who='nom_personnage']/*:p/*:name/  
*:c/text()
```

- 82 fichier txt (propos prononcés par les personnages, sans didascalies)

Personnage et textométrie

Titre	Tokens	Word Types
Les Dehors trompeurs	17330	2871
Ésope à la cour	20643	3220
La Galerie du Palais	16001	2615
Le Chevalier à la mode	20809	2478
Le Glorieux	19048	2944
La Surprise de l'amour	16662	2480
L'École des femmes	15288	2666
Le Joueur	14991	2676

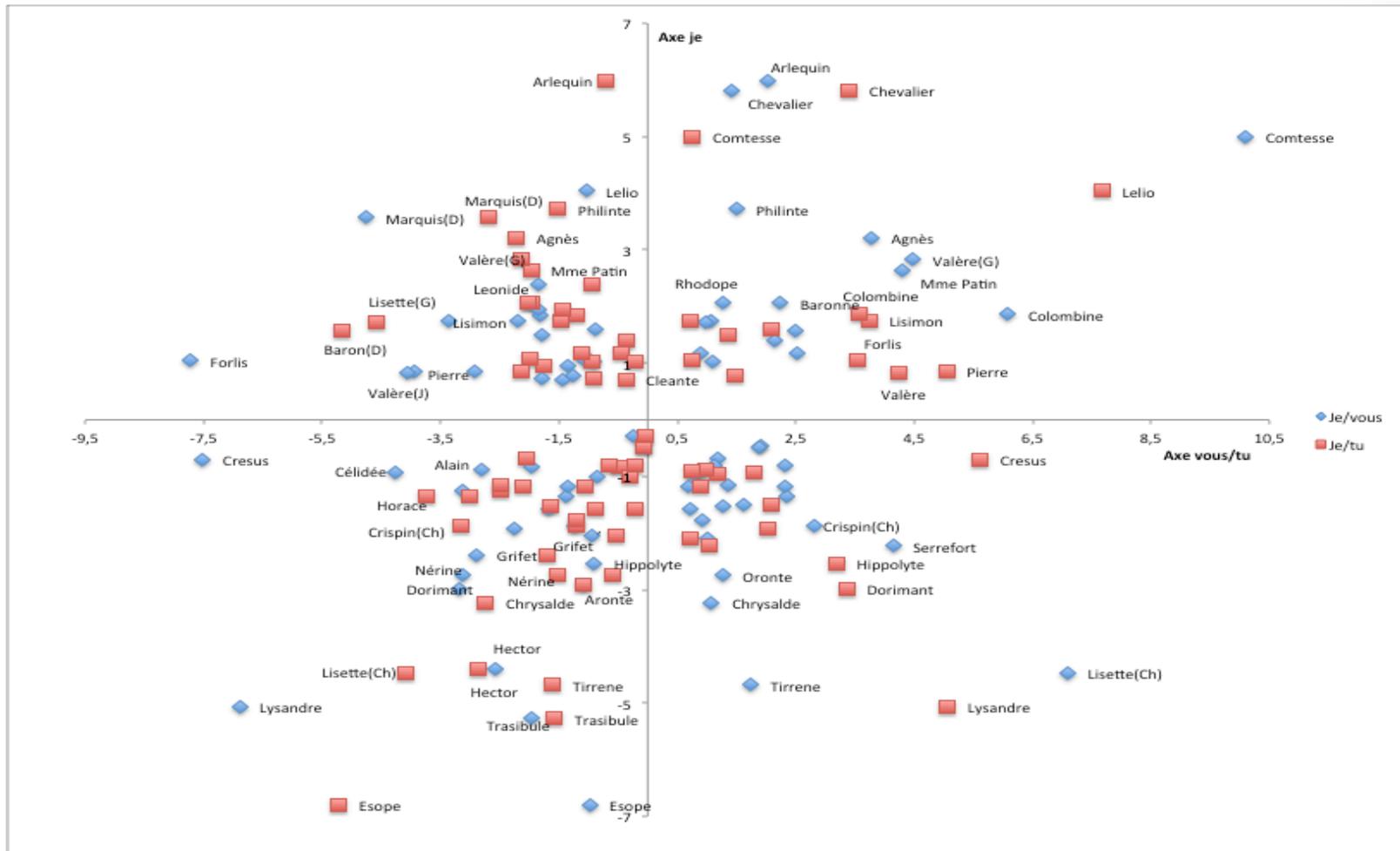
Personnage et textométrie

- Utilisation du logiciel Hyperbase (E. Brunet, Nice)
- « Je », « vous » et « tu » chez les 82 personnages du corpus: « indice de spécificité » (valeurs significatives au-delà de +2/ -2)
- Il existe bien des personnages caractérisés par un usage plus intensif (ou significativement très réduit) des pronoms du dialogue

Personnage et textométrie

- Hyperbase fournit un tableau des écarts réduits, personnage par personnage > possibilité de les projeter sur un graphique
- Quatre catégories de personnages
 1. beaucoup de « je » et peu de « vous » ;
 2. beaucoup de « vous » et peu de « je » ;
 3. à la fois beaucoup de « je » et beaucoup de « vous » ;
 4. « déficit » sur l'un et sur l'autre vocable.
- Idem pour « je »/ « tu »

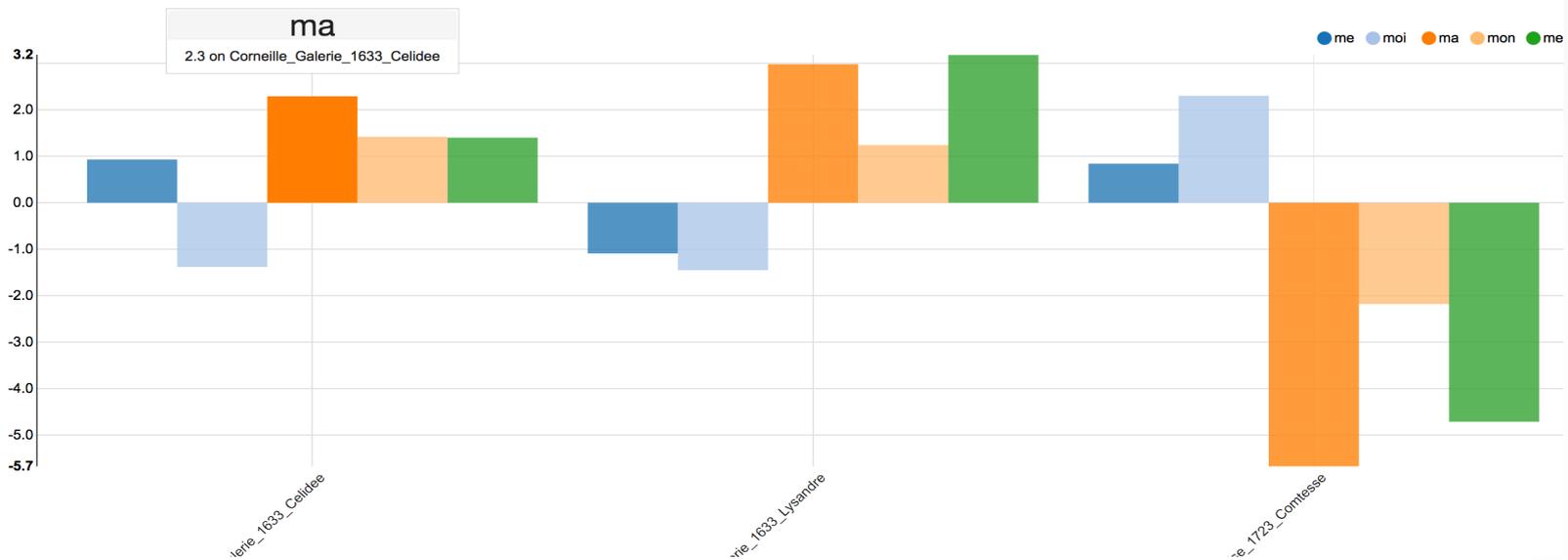
Personnage et textométrie



Personnages et textométrie

- Personnages de Destouches, Marivaux, Boissy en haut du graphique, ceux de Corneille et Molière en bas (sauf Agnès), Dancourt et Regnard en position intermédiaire
- Mais « je » est une mesure incomplète quid de « mon coeur », « ma valeur », « mon... », « ma... »?

Personnages et textométrie



- besoin d'un indice composite
- « je » comme pis aller

Personnages et textométrie

- Que nous disent ces positions sur le plan de la dramaturgie?
- 4 positions significatives
 - a) Quart haut à gauche: personnages « égoïstes ». L'amant (Lisimon du *Glorieux*), le personnage imbu de soi (le Marquis des *Dehors trompeurs*), le personnage d'autorité (Forlis dans DT)

Personnages et textométrie

b) Bas gauche: personnages détachés

- Les valets observateurs: Nérine, Hector
- Les porteurs d'une forme de sagesse et d'un regard en surplomb: Ésope, Célidée

c) Bas droite: des « altruistes »?

- Lisette et Crispin (Le Chevalier à la mode): pas de dessein propre
- mais aussi Serrefort (Chevalier), Hippolyte, Dorimant, Lisandre (Corneille)

➔ Une case pour les personnages « adjuvants » et « opposants »

Personnages et textométrie

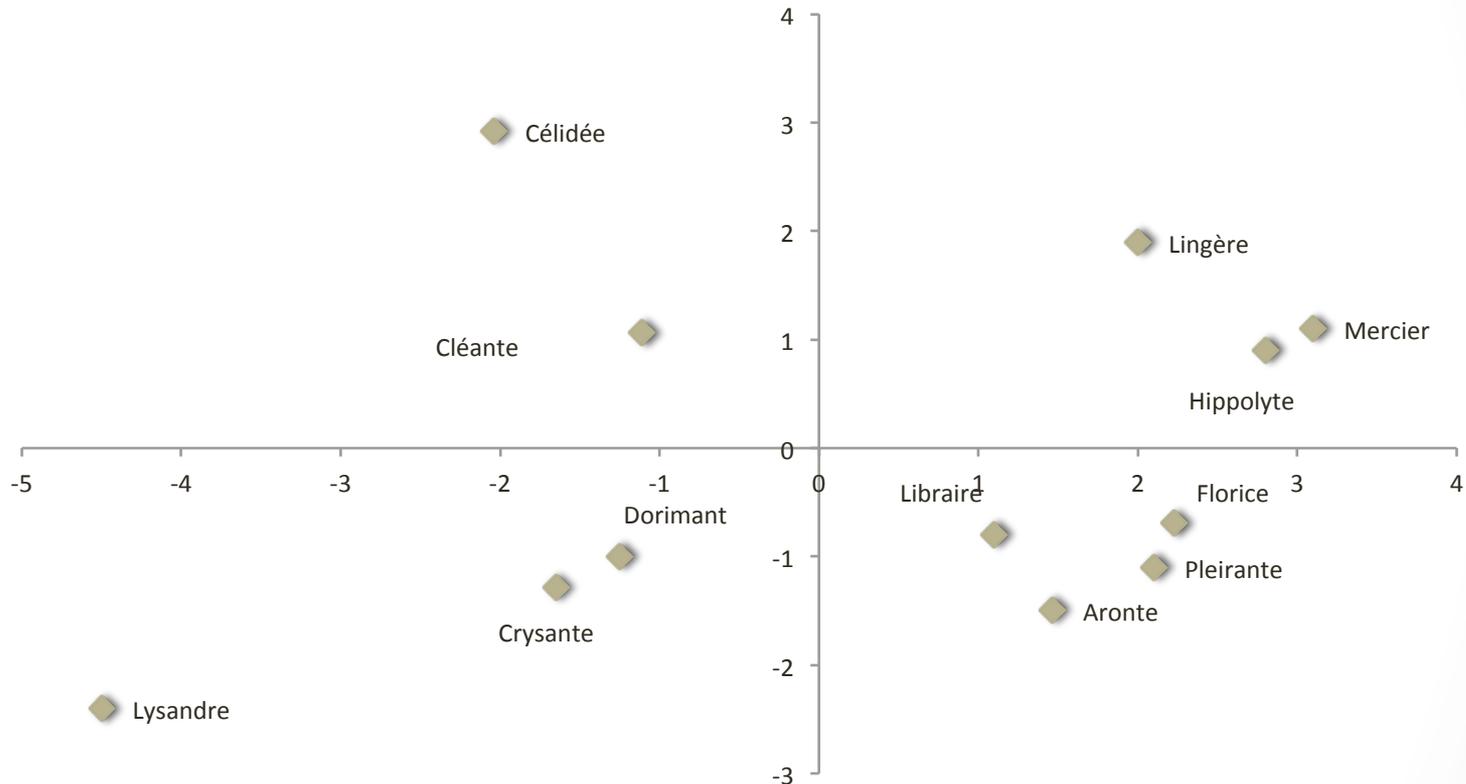
d) Haut droite (beaucoup de « je », beaucoup de « vous »/ « tu »)

- Un sentiment fort, mais personnages attentifs aux désirs et vœux des autres: Agnès (Molière), Valère (Destouches), Arlequin et Colombine (Marivaux), Mme Patin (Dancourt et Saint-Yon)
- Quid du chevalier de Villefontaine?

Personnages et textométrie

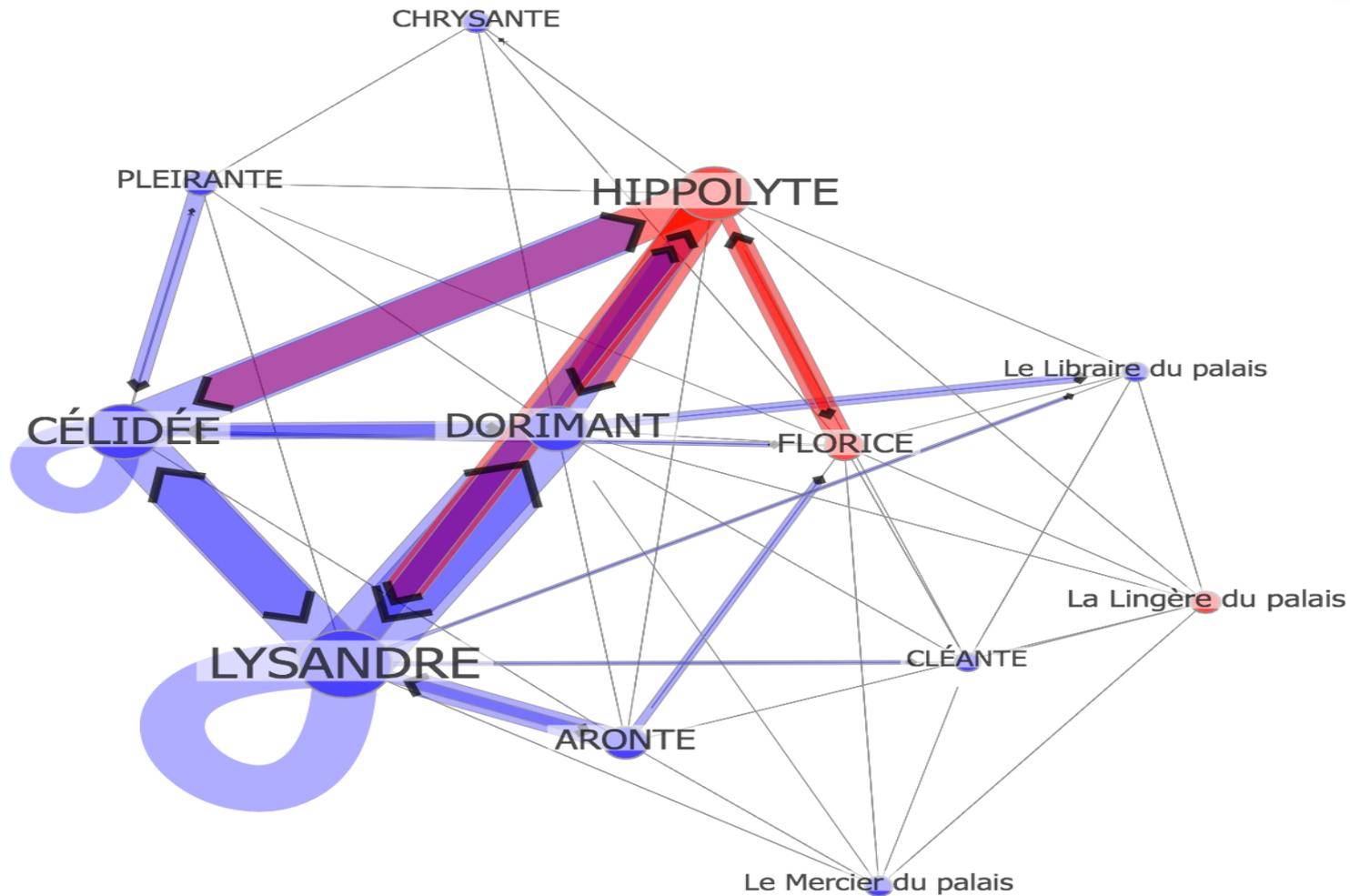
- Outliers (naturels dans toute enquête statistique), ou limites de la comparaison de pièces disparates?
- Normes stylistiques fort différentes entre le temps de Corneille et celui de Boissy
- Même type d'analyse dans des univers homogènes: *La Galerie du Palais* et *Les Dehors trompeurs*

Personnages et textométrie

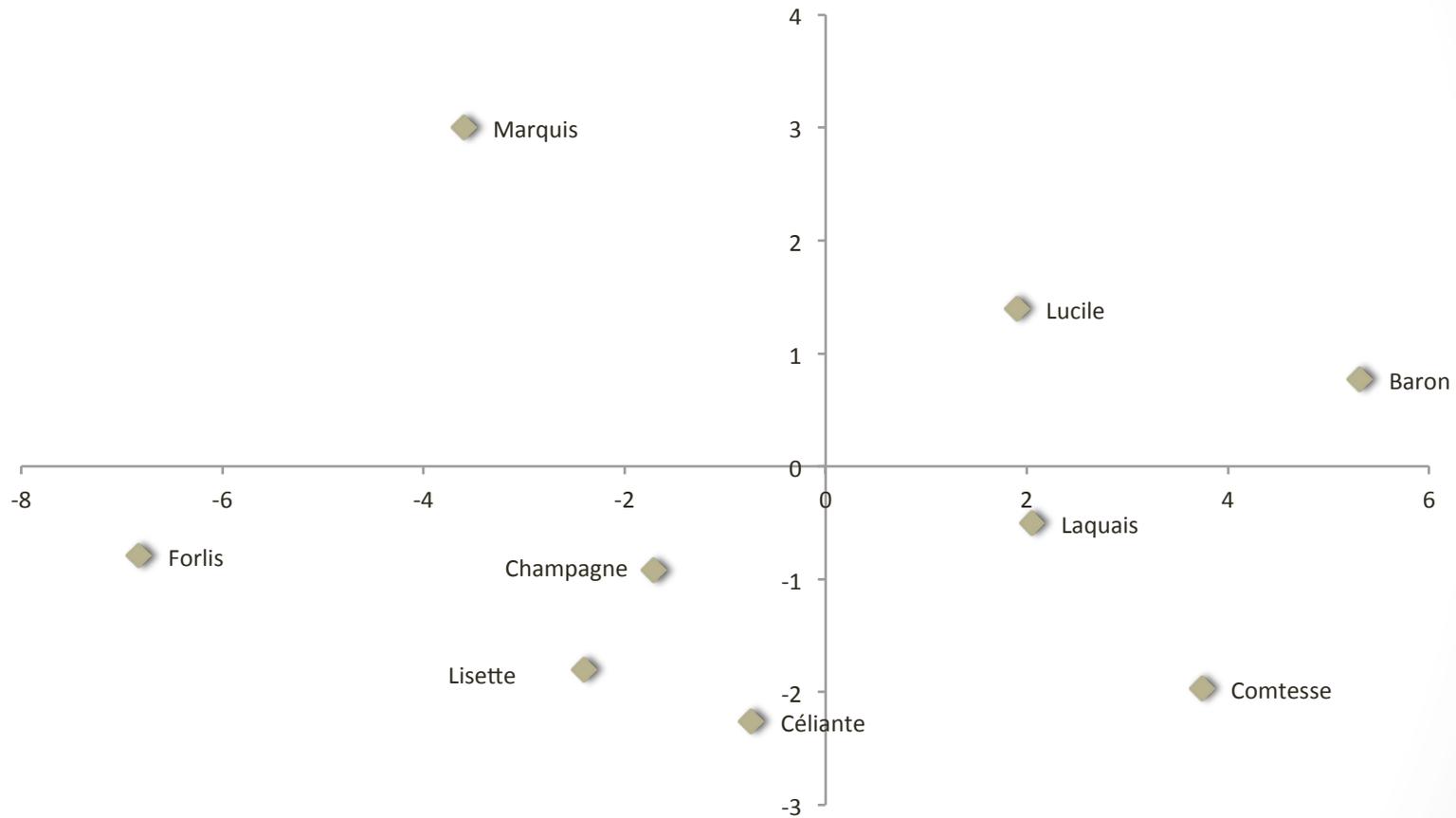


La Galerie du Palais

Personnages et textométrie

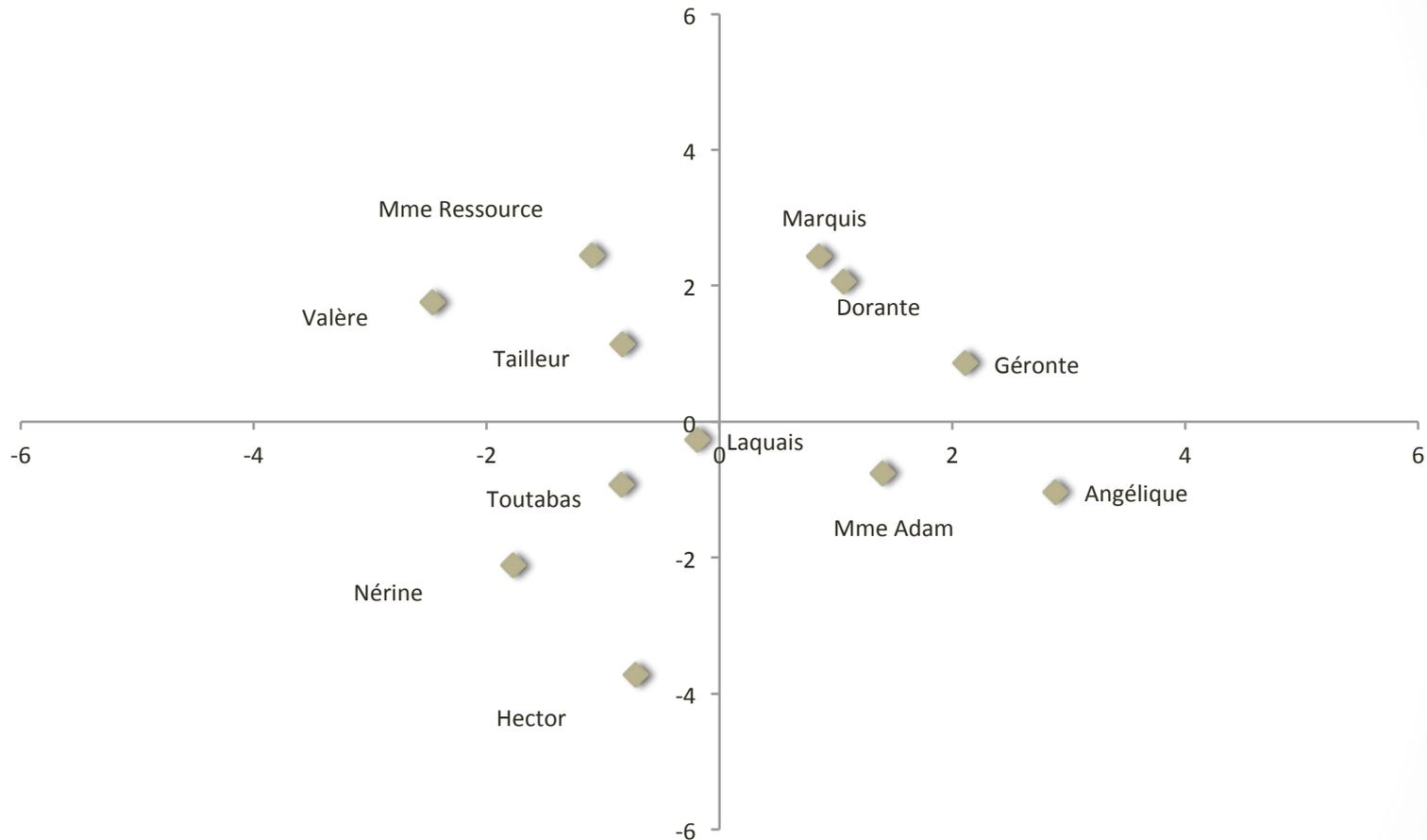


Personnages et textométrie



Les Dehors trompeurs

Personnages et textométrie



Le Joueur

Expériences de balisage

- Un travail bien moins avancé que celui réalisé avec AntConc
- L'orientation de la TEI est traditionnellement plutôt éditoriale, plus récemment philologique, moins analytique
- Pas de balise dédiée pour les personnages: <castItem> ou <speaker> au théâtre, mais pas clair que faire de
« Je viens de chez Clorinde. »
<p>Je viens de chez <name>Clorinde</name></p>

Expériences de balisage

- Au-delà des noms:
 - chaînes de co-références, incluant pronoms et périphrases (« cette belle »)
 - éléments de caractérisation
Ex. Admète, roi de Thessalie
Elle a de babiller une fureur extrême.
 - le personnage sujet et objet de verbes d'action, volition, cognition...

Expériences de balisage

- Quoi baliser?
- Comment baliser?
 - Choix de balises, typage, meilleures pratiques
- Pourquoi baliser?
 - le relevé complet des éléments a-t-il une plus value pour la compréhension du personnage?

Expériences de balisage

- Quoi: une question à double fond
 - quelles entités?
 - quels éléments construisent ces entités?
- L'expérience Cortazar
 - I. Galleron, F. Idmhand, C. Meynard « Que mille lectures s'épanouissent... », *DHQ*, numéro francophone, automne 2017.
 - Une centaine de répondants
 - Entre 3 et 8 personnages relevés

Expériences de balisage

- Une question de « seuil », qui peut être lui-même un objet de recherche
- « Ratissage large »: tout ce qui fait référence à des figures anthropomorphes dans les pièces de théâtre
- « tout » = noms, syntagmes nominaux, attributs du sujet, appositions + pronoms (?); problème des possessifs
- relevé des éléments de persistance et de description, pas d'action/ volition/ cognition

Expériences de balisage: comment?

- Balises `<castItem>`, `<name>` et `<rs>`, complétées de trois attributs
- `@ref` pointant vers un `xml:id` attribué au personnage dans la `castList` ou dans une liste complémentaire inscrite dans `<front>`
- `@type` sur `<name>`, indiquant l'appartenance au monde des caractères (`type='character'`)
- `@ana` sur `<rs>`, concaténant des codes de caractérisation

Expériences de balisage: comment?

- Génération d'une deuxième liste de personnages dans le header

```
<item xml:id="Marquis" ana="#OCH #SM #AGY #ARL #ARS #SS #FPB">  
  <rs>Le Marquis, amant de Paméla.</rs> </item>
```

```
<item xml:id="Pamela" ana="#OCH #SF #AGY #ARL #FPD">  
  <name>Paméla</name> </item>
```

```
<item xml:id="Chevalier" ana="#OCH #SM #ARL #AROP #SS #NAT">  
  <rs>Le Chevalier, Gascon.</rs></item>
```

```
<item xml:id="Nerine" ana="#OCH #SF #AGU #ARAD #SS #OCC">  
  <name>Nérine, concierge.</name></item>
```

```
<item xml:id="Mathurin" ana="#OCH #SM #AGU #ARL #ARS #SS  
#OCC">
```

```
  <name>MATHURIN, jardinier.</name> </item>
```

```
<item xml:id="Jeukes" ana="#OCH #SM #AGU">  
  <name>Maître Jeukes</name> </item>
```

Expériences de balisage: comment?

```
- <item xml:id="Lutin1" ana="#OCNH #SU #COLL #AROP #ARAD">  
    <rs>Premier lutin</rs>  
</item>  
<item xml:id="Lutin2" ana="#OCNH #SU #COLL #AROP #ARAD">  
    <rs>Deuxième lutin</rs>  
</item>  
<item xml:id="Danseurs" ana="#COLL #OCC">  
    <rs>Danseurs</rs>  
</item>  
<item xml:id="Pere" ana="#OCH #SM #AGU #FPF">  
    <rs>Père de Paméla</rs>  
</item>  
<item xml:id="Mere" ana="#OCH #SF #AGU #FPM">  
    <rs>Mère de Paméla</rs>  
</item>  
<item xml:id="Maître" ana="#OCH #SM #AGU #SS">  
    <rs>Ancien maître de Paméla.</rs>  
</item>  
<item xml:id="Comtesse" ana="#OCH #SF #AGU #SS #FPSIS">  
    <rs>La Comtesse</rs>  
</item>
```

Expériences de balisage: une typologie de caractérisation

I. Galleron « Conceptualisation of theatrical characters in the digital paradigm: needs, problems and foreseen solutions », *Human and Social studies*, De Gruyter, vol. 6, issue 1 (**Published Online: 2017-04-18 | DOI: <https://doi.org/10.1515/hssr-2017-0007>**)

- Statut ontologique du personnage: OCH/ OCNH/ OCSH
- Sexe: SM/ SF/ SU
- Âge: AGY/ AGO/ AGC/ AGA/ AGU
- Nationalité, origine: NAT

Expériences de balisage: codes de caractérisation

- Religion, croyances: REL
- Opinions politiques, positionnement: POL
- Affinités, préférences, caractéristiques culturelles: CULT
- Statut social: SS
- Occupation: OCC
- Personnage collectif: COLL

Expériences de balisage: codes de caractérisation

- Position familiale: FPH, FPWIF, FPF, FPM, FPSON, FPD, FPA, FPU, FPN, FPG, FPWAR, FPWID, FPWDR, FPC, FPB, FPSIS, FPNEIG, FPFRI
- Rôle actantiel: ARS, ARL, ARAUT, ARAD, AROP, AREP, ARTM
- Aspect physique: PHYS
- Caractéristiques morales: TEMP

Problèmes de codage

- `<l><rs ref="#Nerine" ana="#TEMP">Je laiffe <rs ref="#Marquis">le Marquis</rs>, & n'ai pas la conftance</rs>.</l>`
- `<l part="F"><rs ref="#Nerine">Je</rs> dis malgré la réfiftance</l>
<l>Que <rs ref="#Pamela" ana="#ARL">cette fille l'aime, & qu'elle reftera</rs>.</l>`
- `<l><rs ref="#Pamela" type="MET">Son cœur</rs>, de fon départ, le confie à <rs ref="#Chevalier" type="MET" ana="#TEMP">ma foi</rs> ;</l>`

Problèmes de codage

- `<l part="l">Sur ma parole....</l> </sp>`
`<sp xml:id='ll_9_15' who="#Nerine">`
`<speaker>NERINE<pc type="affirmative">.</pc></speaker>`
- `<l>Bon <rs ref="#Chevalier" ana="#TEMP">vous n'en avez`
`aucune</rs>.</l>`
- Là, crac, `<rs ref="#Chevalier">j</rs>'enleve <rs`
`ref="#Pamela">mon Europe</rs>`
- `<rs ref="#Chevalier">Nous</rs> fommes vrais tous deux`

Expériences de balisage: quels résultats?

- Utilisation de BaseX pour effectuer des extractions
- Corpus Boissy: 37 pièces
- 280 personnages (<castItem>)

174 OCH

21 OCNH

10CSH

Expériences de balisage: quels résultats?

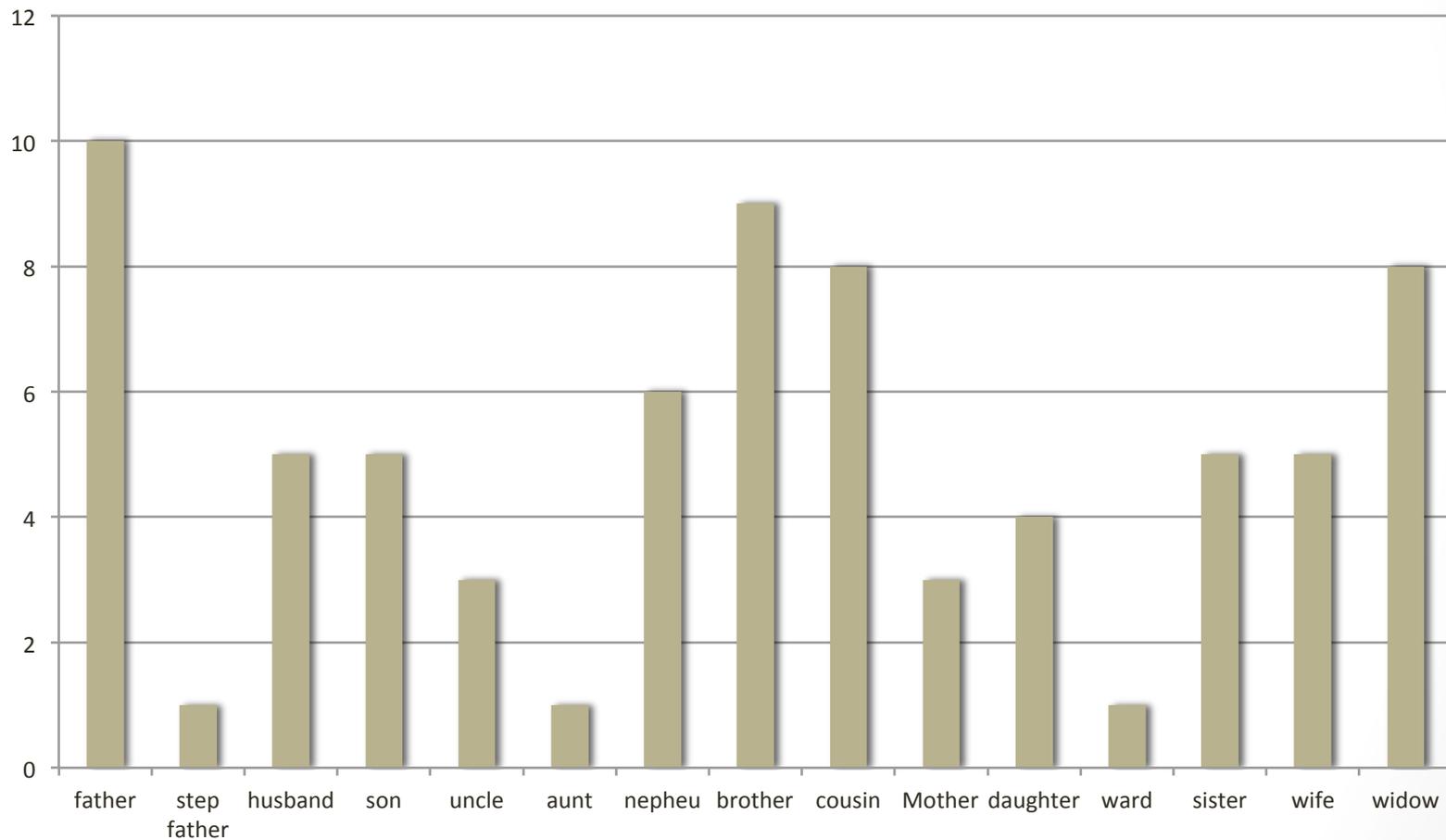
- Sexe: 119 M, 67 F, 10 autres (SU)
- Âge: 43 AGY, 11 AGO, 21 AGA, 1 AGC, 77 AGU
- 177 actantial roles (mais nombreux personnages sans rôle actantiel, même confus ex. « troupe de danseurs »)
- 52 ARL, 42 ARS, 34 ARAD, 27 AROP, 54 AREP

Expériences de balisage: quels résultats?

- 9 personnages collectifs
- seul un personnage sur trois caractérisé par la position familiale: 60 FP

Est-ce beaucoup? Peu? Avec quoi comparer?

Expériences de balisage: quels résultats?



Expériences de balisage: quels résultats?

L'expérience *Paméla en France*

- 1366 « reference strings »
- 288 rs porteurs d'une caractérisation (ana)
 - 90 concernent Paméla
 - 25 concernant Le Marquis
 - 26 concernant Le Chevalier
 - 9 Nérine
 - 8 Comtesse (personnage absent)

Expériences de balisage: quels résultats?

- 117 TEMP
- 26 SS
- 22 PHYS
- 18 FP
- 12 OCC
- 11 CULT
- 7 NAT

Expériences de balisage: quels résultats?

Pamela_en_France.xml
TEI

teiHeader	text
fileDesc	body
titleS... n.. sour..	div
pub... n.. bibl..	div
encodingDesc	div
projectDesc	div
p p p	sp s.. s.. sp sp s..
p p p	sp s..
p p p	sp s.. s.. sp sp s..
p p p	sp s.. s.. sp sp s..
editorialDecl	sp s.. s.. sp sp s..
n.. h.. interp..	sp s.. s.. sp sp s..
fsdDecl	sp s.. s.. sp sp s..
fsDecl	sp s.. s.. sp sp s..
family_..	div
age	div
vRange	sp s.. s.. sp sp s..
vColl	sp s.. s.. sp sp s..
s..	sp s.. s.. sp sp s..
ont..	sp s.. s.. sp sp s..
actantial_.. c..	sp s.. s.. sp sp s..
profileDesc	sp s.. s.. sp sp s..
particDesc	sp s.. s.. sp sp s..
p	sp s.. s.. sp sp s..
list	sp s.. s.. sp sp s..

Merci de votre attention!

ioana.galleron@univ-paris3.fr