

Analyse et extraction des motifs syntaxiques dans la prose de Robert Challe et de ses apocryphes

Séminaire OBVIL Challe - Paris, le 28 mars 2015

Francesca Frontini**, Amine Boukhaled, Jean-Gabriel Ganascia
Laboratoire d'informatique de Paris 6 -Labex OBVIL

**Boursière Fernand Braudel de la *Fondation Maison Sciences de l'homme* - Paris



Stylistique computationnelle

Domaine de recherche qui étudie les méthodes automatiques d'identification ou d'extraction de motifs de style qui caractérisent un type particulier de textes.

- ❑ Type particulier de textes :
 - ❑ même genre ,
 - ❑ écrits par le même auteur
- ❑ Deux types de phénomènes stylistiques intéressants à étudier :
 - ❑ **la déviation**
 - ❑ la répétition

Notre approche pour l'étude de la déviation

Objectifs:

- 1) Identifier les similarités entre les textes
- 2) Identifier les traits stylistiques importants pour chaque texte

Méthodologie:

Analyse statistique et visualisation afin de déterminer les séquences syntagmatiques les plus typiques et significatives pour chaque texte.

Corpus - *Les Illustres Françaises*

Authentiques:

- Histoire de Monsieur de **Contamine** et d'Angélique.
- Histoire de Monsieur **Dupuis** et de Madame de Londé.
- Histoire de Monsieur Des **Frans** et de Silvie.
- Histoire de Monsieur de **Jussy** et de Mademoiselle Fenouil.
- Histoire de Monsieur Des **Prez** et de Mademoiselle de l'Épine.
- Histoire de Monsieur Des **Ronais** et de Mademoiselle Dupuis.
- Histoire de Monsieur de **Terny** et de Mademoiselle de Bernay.

Sans introduction.

Apocryphes:

- Histoire de Monsieur de **Bréville** et de Mademoiselle de Beaumont.
- Histoire de Madame **Darmentin**.
- Histoire du comte de **Livry** et de Mademoiselle de Mancigni.
- Histoire de Monsieur de **Salvagne** et de Madame de Villiers.
- Histoire de Monsieur le comte de **Vallebois** et de Mademoiselle Charlotte de Pontais, son épouse.

Méthodologie

Annotation automatique du texte en classes syntagmatiques

Le livre est sur la table .

DET NOM VER PRP DET NOM SENT

Extraction non supervisée de motifs syntaxiques de taille 3,4 ou 5

DET NOM VER PRP DET NOM SENT

Visualisation et extraction des motifs importants avec la méthodologie statistique de l'analyse des correspondances

Méthodologie

Annotation automatique du texte en classes syntagmatiques

Le livre est sur la table .

DET NOM VER PRP DET NOM SENT

Extraction non supervisée de motifs syntaxiques de taille 3,4,5

DET NOM VER PRP DET NOM SENT

Visualisation et extraction des motifs importants avec la méthodologie statistique de l'analyse des correspondances

Méthodologie

Annotation automatique du texte en classes syntagmatiques

Le livre est sur la table .

DET NOM VER PRP DET NOM SENT

Extraction non supervisée de motifs syntaxiques de taille 3,4,5

DET **NOM VER PRP DET NOM** SENT

Visualisation et extraction des motifs importants avec la méthodologie statistique de l'analyse des correspondances

Méthodologie

Annotation automatique du texte en classes syntagmatiques

Le livre est sur la table .

DET NOM VER PRP DET NOM SENT

Extraction non supervisée de motifs syntaxiques de taille 3,4,5

DET NOM VER PRP DET NOM SENT

Visualisation et extraction des motifs importants avec la méthodologie statistique de l'analyse des correspondances

Méthodologie

Annotation automatique du texte en classes syntagmatiques

Le livre est sur la table .

DET NOM VER PRP DET NOM SENT

Extraction non supervisée de motifs syntaxiques de taille 3,4,5

DET **NOM VER PRP** DET NOM SENT

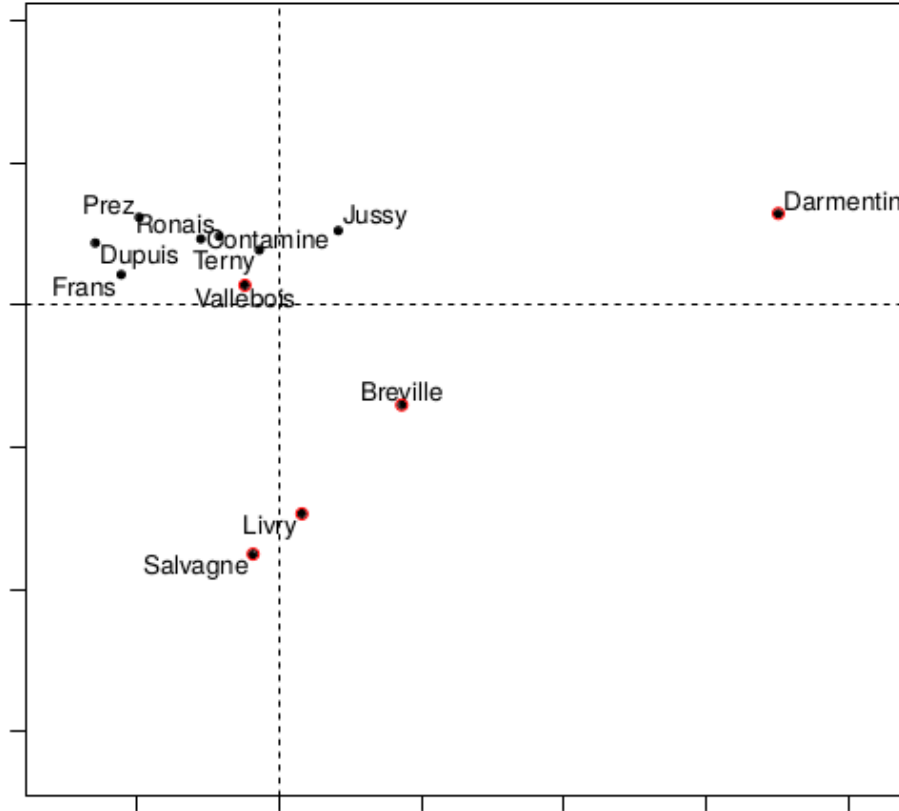
Visualisation et extraction des motifs importants avec la méthodologie statistique de l'analyse des correspondances

La méthode d'analyse des motifs

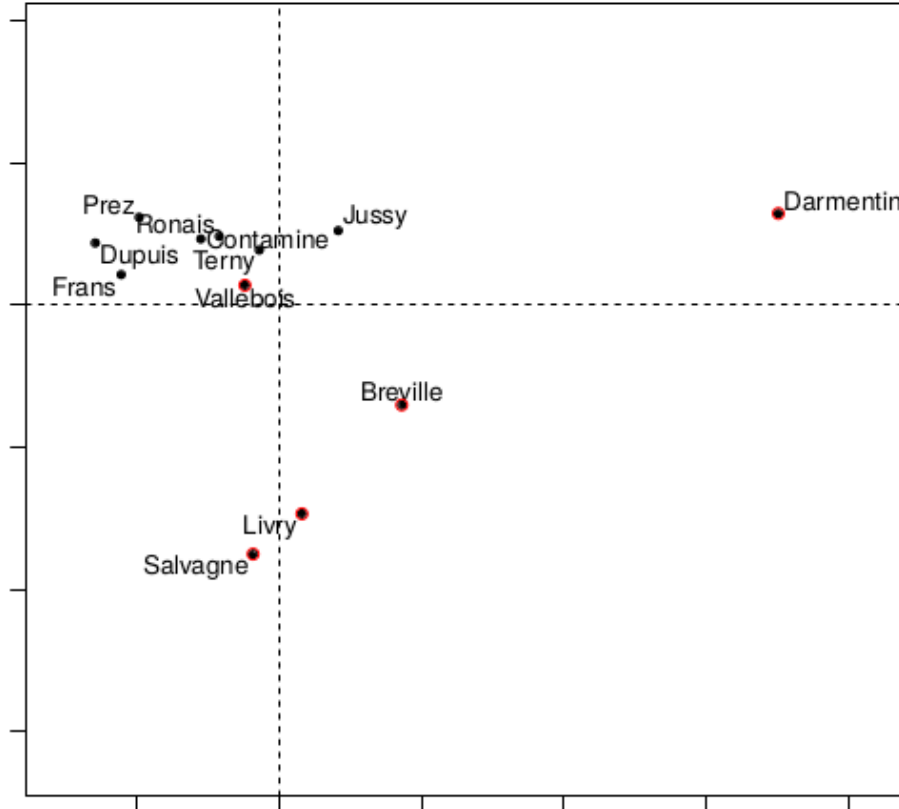
Analyse des correspondances (Benzécri, 1977): technique d'analyse statistique multivariée.

- ❑ Projection des textes et de leurs motifs dans un espace bi-dimensionnel.
- ❑ Classement des motifs en fonction de leur contribution combinée sur les deux axes, en permettant ainsi au chercheur de filtrer les motifs les moins intéressants, c'est-à-dire ceux qui ont une faible contribution.

Projection des textes

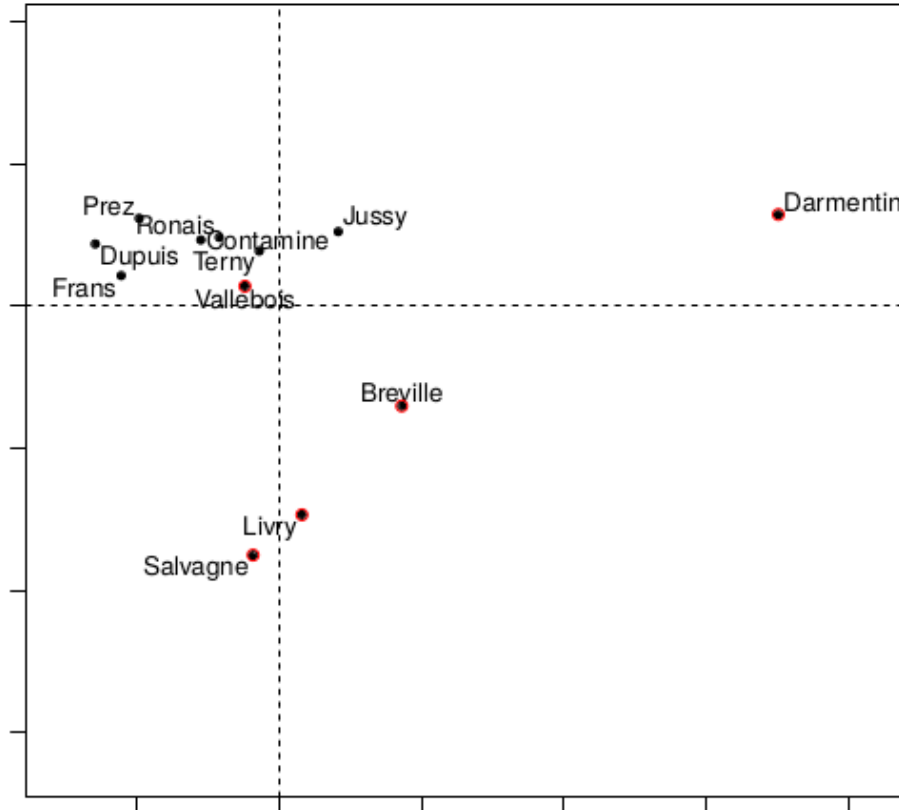


Projection des textes



Ce tableau représente sous forme de **distances** les **différences** entre les textes par rapport aux **fréquences** des motifs syntaxiques extraits.

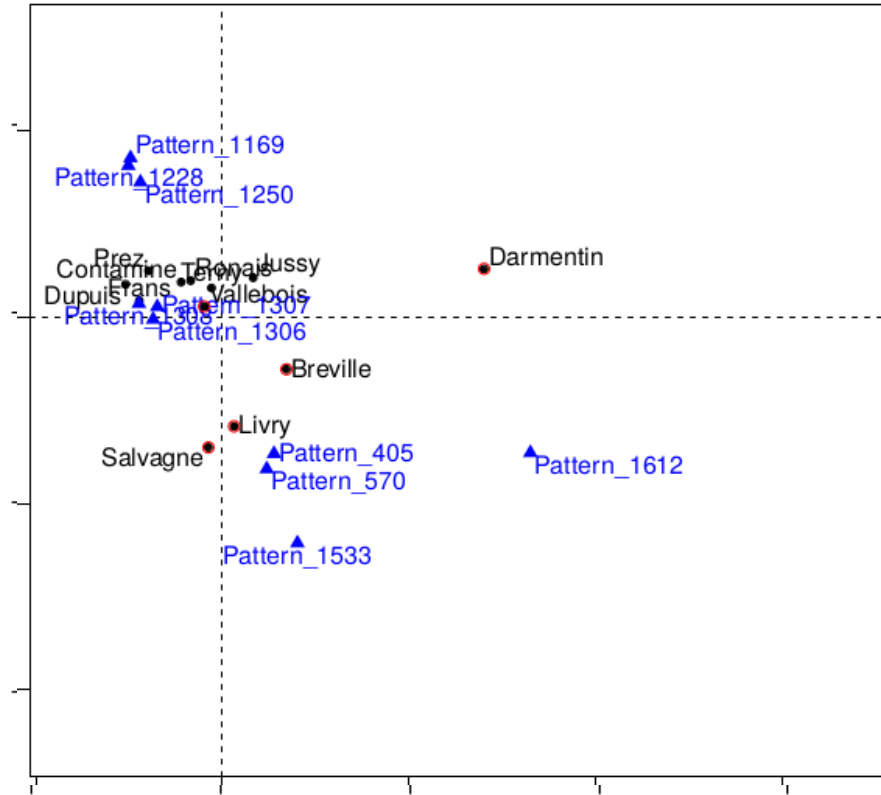
Projection des textes



Ce tableau représente avec des **distances** les **différences** entre les textes par rapport aux **fréquences** des motifs syntaxiques extraits.

1. Les apocryphes (sauf Vallebois) semblent avoir un comportement syntaxique différent des authentiques.
2. Les originaux sont très proches les uns des autres.
3. Les apocryphes sont différents les uns des autres.

Projection des 10 motifs le plus importants

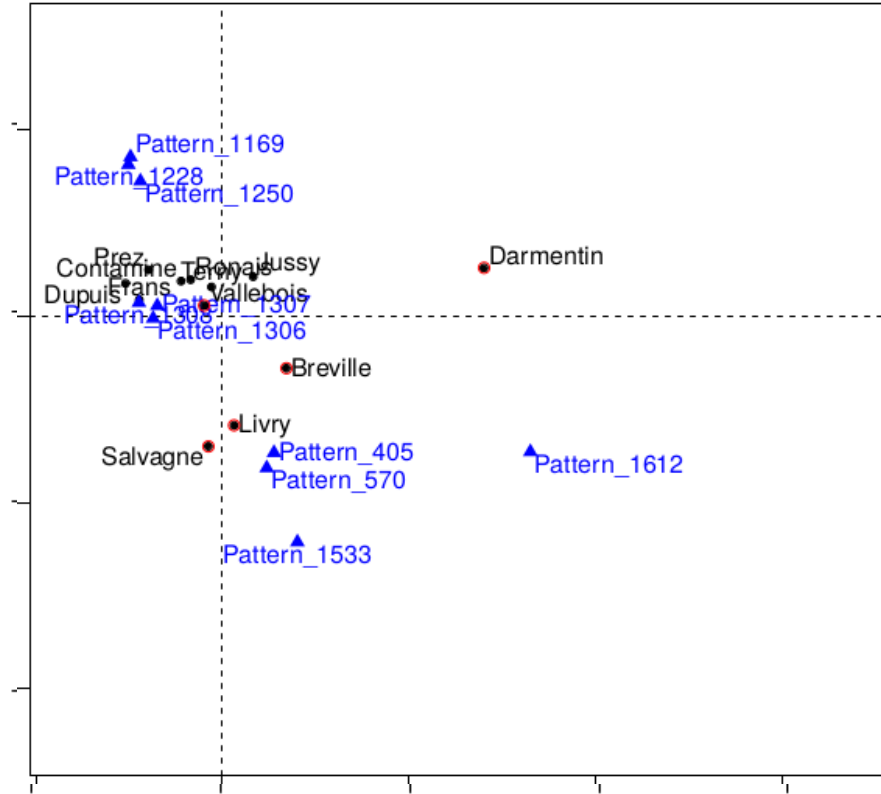


Projection des 10 motifs le plus importants

Le codes des motifs sont représentés dans l'espace, **près** des textes dans lesquels ils sont plus fréquents.

Motif important = motif dont la fréquence est particulièrement hétérogène dans les textes.

Les motifs les plus importants contribuent à l'**éloignement** des textes sur les deux axes.

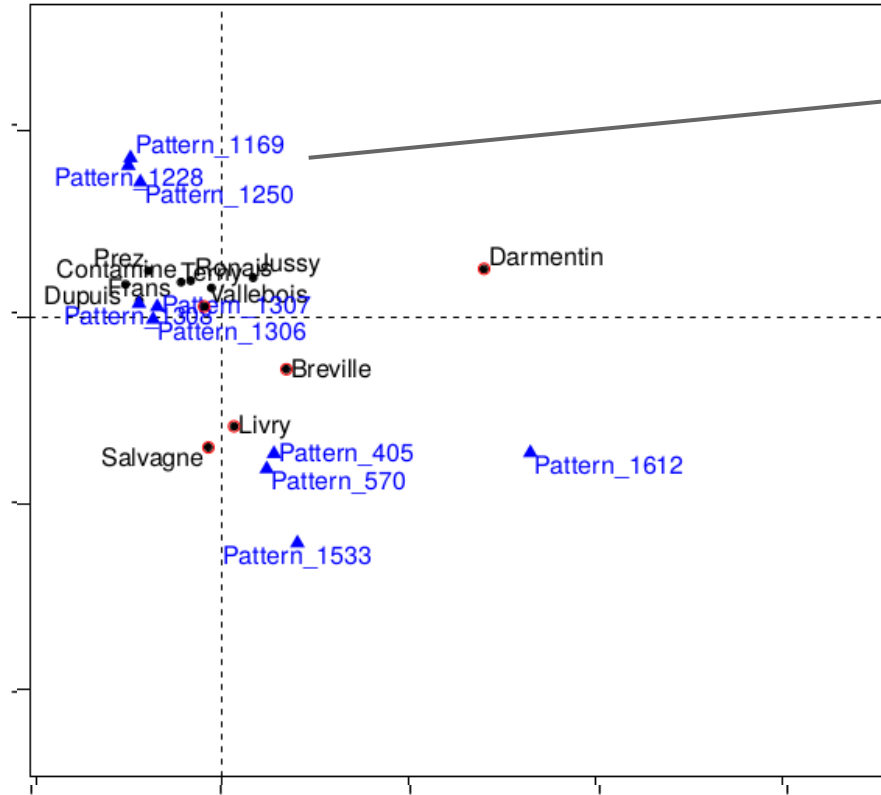


Projection des 10 motifs le plus importants

Le codes des motifs sont représentés dans l'espace, **près** des textes dans lesquels ils sont plus fréquents.

Motif important = motif dont la fréquence est particulièrement hétérogène dans les textes.

Les motifs les plus importants contribuent à l'**éloignement** des textes sur les deux axes.



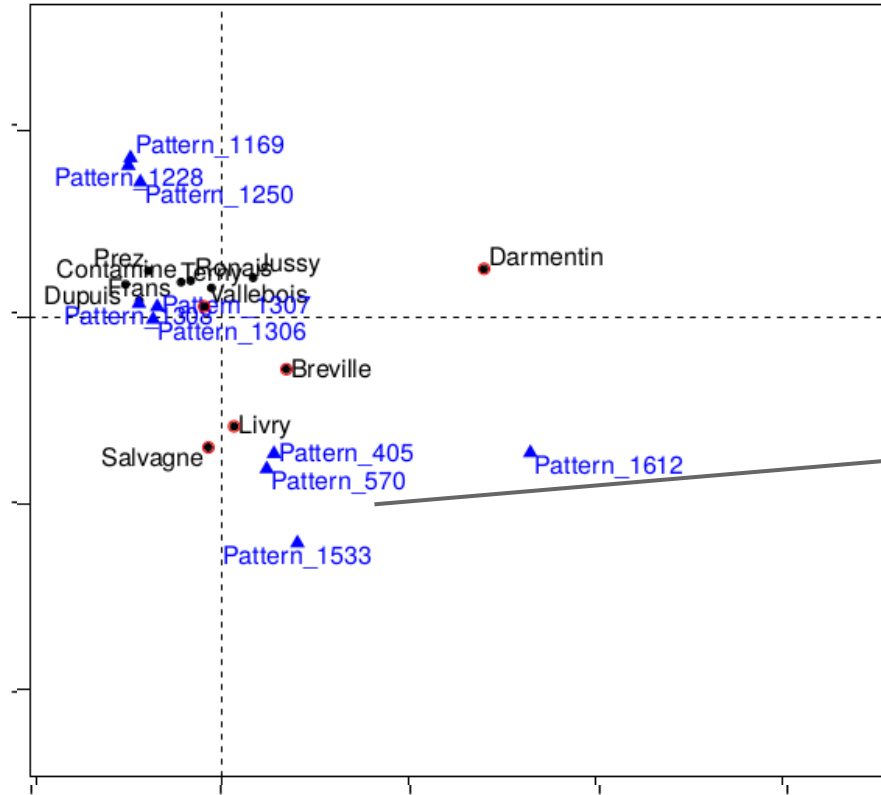
Motifs plus associés aux authentiques.

Projection des 10 motifs le plus importants

Le codes des motifs sont représentés dans l'espace, **près** des textes dans lesquels ils sont plus fréquents.

Motif important = motif dont la fréquence est particulièrement hétérogène dans les textes.

Les motifs les plus importants contribuent à l'**éloignement** des textes sur les deux axes.



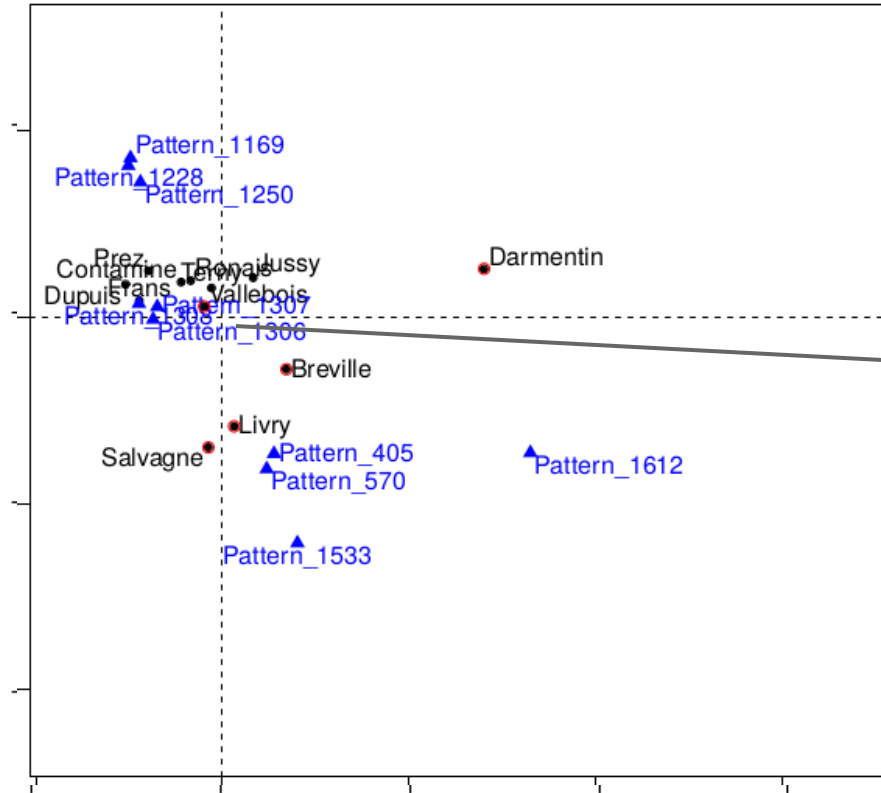
Motifs plus associés aux apocryphes.

Projection des 10 motifs le plus importants

Le codes des motifs sont représentés dans l'espace, **près** des textes dans lesquels ils sont plus fréquents.

Motif important = motif dont la fréquence est particulièrement hétérogène dans les textes.




Les motifs les plus importants contribuent à l'**éloignement** des textes sur les deux axes.



Motifs communs entre Vallebois et les authentiques.




Analyse des motifs

ID du motif	Motif	Texte le plus proche
Pattern_1228	PUN - VER - PRO	PREZ
Pattern_1308	PRO - PRO - VER	FRANS
Pattern_1250	VER - PRO - PUN	PREZ
Pattern_1533	PRP - NAM - PPR - NAM	SALVAGNE
Pattern_1169	PUN - VER - PRO - PUN	PREZ
Pattern_570	NAM - PRP - NAM	LIVRY
Pattern_1307	PRP - DET - NOM	FRANS
Pattern_405	PRP - NAM - PRP	LIVRY
Pattern_1306	PRO - VER - PRP	FRANS
Pattern_1612	VER - NOM - PRP - NAM	DARMENTIN

-  proche des authentiques
-  proche des authentiques + Vallebois
-  proche des apocryphes




Analyse des motifs

ID du motif	Motif	Texte le plus proche
Pattern_1228	PUN - VER - PRO	PREZ
Pattern_1308	PRO - PRO - VER	FRANS
Pattern_1250	VER - PRO - PUN	PREZ
Pattern_1533	PRP - NAM - PPR - NAM	SALVAGNE
Pattern_1169	PUN - VER - PRO - PUN	PREZ
Pattern_570	NAM - PRP - NAM	LIVRY
Pattern_1307	PRP - DET - NOM	FRANS
Pattern_405	PRP - NAM - PRP	LIVRY
Pattern_1306	PRO - VER - PRP	FRANS
Pattern_1612	VER - NOM - PRP - NAM	DARMENTIN

-  proche des authentiques
-  proche des authentiques + Vallebois
-  proche des apocryphes




Analyse des motifs

ID du motif	Motif	Texte le plus proche
Pattern_1228	PUN - VER - PRO	PREZ
Pattern_1308	PRO - PRO - VER	FRANS
Pattern_1250	VER - PRO - PUN	PREZ
Pattern_1533	PRP - NAM - PPR - NAM	SALVAGNE
Pattern_1169	PUN - VER - PRO - PUN	PREZ
Pattern_570	NAM - PRP - NAM	LIVRY
Pattern_1307	PRP - DET - NOM	FRANS
Pattern_405	PRP - NAM - PRP	LIVRY
Pattern_1306	PRO - VER - PRP	FRANS
Pattern_1612	VER - NOM - PRP - NAM	DARMENTIN

-  proche des authentiques
-  proche des authentiques + Vallebois
-  proche des apocryphes

Analyse des motifs

ID du motif	Motif	Texte le plus proche
Pattern_1228	PUN - VER - PRO	PREZ
Pattern_1308	PRO - PRO - VER	FRANS
Pattern_1250	VER - PRO - PUN	PREZ
Pattern_1533	PRP - NAM - PPR - NAM	SALVAGNE
Pattern_1169	PUN - VER - PRO - PUN	PREZ
Pattern_570	NAM - PRP - NAM	LIVRY
Pattern_1307	PRP - DET - NOM	FRANS
Pattern_405	PRP - NAM - PRP	LIVRY
Pattern_1306	PRO - VER - PRP	FRANS
Pattern_1612	VER - NOM - PRP - NAM	DARMENTIN

-  proche des authentiques
-  proche des authentiques + Vallebois
-  proche des apocryphes

Notre outil d'extraction de motifs permet aussi d'extraire les instances textuelles pour chaque motif dans les différents textes.

Les résultats se présentent comme des **listes** qui montrent le **contexte** de la phrase, pour permettre une analyse plus détaillée.

Motifs dans les authentiques - [PUN - VER - PRO]

Voici quelques passages dans *Histoire de Monsieur Des Prez, et de Mademoiselle de l'Épine* où l'on trouve la séquence [PUN - VER - PRO]

Instance du motif : , **repris-je**

La phrase : J'en crois connaître le caractère, **repris-je**, et qu'elle est de l'écriture de Mademoiselle de l'Épine .

Instance du motif : , **poursuivis -je**

La phrase : Mais comment peut-elle vous écrire, puisqu'elle est morte, **poursuivis-je**, à ce que tout le monde dit ?

Instance du motif : , **répondit-il**

La phrase : Elle l'est aussi, **répondit-il**, et plutôt à Dieu qu'elle ne la fût pas, je ne serais pas ici, mais elle ne serait pas tout à fait perdue pour moi !

Instance du motif : , **poursuivis-je**

La phrase : Mais, **poursuivis-je** dans ce sentiment, pourquoi la mort de cette fille vous est-elle si sensible, puisque vous l'avez abandonnée pendant sa vie ?

Motifs dans les authentiques - [PUN - VER - PRO]

Voici quelques passages dans *Histoire de Monsieur Des Prez, et de Mademoiselle de l'Épine* où l'on trouve la séquence [PUN - VER - PRO]

Instance du motif : , **repris-je**

La phrase : J'en crois connaître le caractère, **repris-je**, et qu'elle est de l'écriture de Mademoi

Instance du motif : , **poursuivis -je**

La phrase : Mais comment peut-elle vous écrire, puisqu'elle est morte, **poursuivis-je**, à ce qu

Instance du motif : , **répondit-il**

La phrase : Elle l'est aussi, **répondit-il**, et plutôt à Dieu qu'elle ne la fût pas, je ne serais pas ici, moi !

Instance du motif : , **poursuivis-je**

La phrase : Mais, **poursuivis-je** dans ce sentiment, pourquoi la mort de cette fille vous est-elle si sensible, puisque vous l'avez abandonnée pendant sa vie ?

Dans tous les textes, les apocryphes comme les authentiques, cette structure semble être utilisée surtout dans le dialogue.

Cependant, cette structure est plus fréquemment utilisée dans les authentiques.

C'est l'indice d'une narration en première personne.

Motifs dans les apocryphes - [VER - NOM - PRP - NAM]

Voici quelques passages dans *Histoire de Madame Darmentin* où l'on trouve la séquence [VER - NOM - PRP - NAM]

Instance du motif : reprit Monsieur de Bréville

La phrase : – Je le sais depuis longtemps, **reprit Monsieur de Bréville**, en soupirant ; mais mes oncles, n'avez-vous rien à m'en dire ?

Instance du motif : repartit Monsieur de St

La phrase : – Vous en parlez là dans des termes un peu forts, **repartit Monsieur de St. Eugène**

Instance du motif : demanda Monsieur de Bréville

La phrase : – Et où demeure-t-elle à présent, lui **demanda Monsieur de Bréville** ?

Instance du motif : répondit Monsieur de St

La phrase : – Elle demeure proche les Incurables, lui répondit **Monsieur de St. Eugène** .

On compare avec ce motif des apocryphes, qui indique une narration en troisième personne.

Motifs dans les apocryphes - [PRP - NAM - PPR - NAM]

Voici quelques passages dans *Histoire de Monsieur de Salvagne et de Madame de Villiers* où l'on trouve la séquence [PRP - NAM - PPR - NAM]

Instance du motif : chez Madame de Villiers

La phrase : Un jour que j' entrai **chez Madame de Villiers**, je trouvai qu'il y était déjà, et qu'il y débitait à son ordinaire quelques lieux communs tirés des romans de Mademoiselle de Scudéry, qu'il avait si bien étudiés, qu'il les savait presque par cœur.

Instance du motif : de Mademoiselle de Scudéry

La phrase : Un jour que j'entrai chez Madame de Villiers, je trouvai qu'il y était déjà, et qu'il y débitait à son ordinaire quelques lieux communs tirés des romans **de Mademoiselle de Scudéry**, qu'il avait si bien étudiés, qu'il les savait presque par cœur.

Instance du motif : pour Mademoiselle de Baugé

La phrase : Madame de Villiers voyant que c'était avec connaissance de cause que j'avais parlé, se mit aussi de la partie, et lui dit qu'elle était surprise qu'il se défendît d'avoir **pour Mademoiselle de Baugé** un attachement, qui ne pouvait que lui faire honneur : qu'elle la connaissait, que c'était une très aimable personne, et qu'elle ne le croyait pas assez malhonnête homme pour la vouloir abuser.

Motifs dans les apocryphes - [PRP - NAM - PPR - NAM]

Voici quelques passages dans *Histoire de Monsieur de Salvagne et de Madame de Villiers* où l'on trouve la séquence [PRP - NAM - PPR - NAM]

Instance du motif : chez Madame de Villiers

La phrase : Un jour que j' entrai **chez Madame de Villiers**, je trouvai qu'il y était déjà, et qu'il y débitait à son ordinaire quelques lieux communs tirés des romans de Mademoiselle de Scudéry, qu'il avait si bien étudiés, qu'il les savait presque par cœur.

Instance du motif : de Mademoiselle de Scudéry

La phrase : Un jour que j'entrai chez Madame de Villiers, je trouvai qu'il y était déjà, et qu'il y débitait à son ordinaire quelques lieux communs tirés des romans **de Mademoiselle de Scudéry**, qu'il avait si bien étudiés, qu'il les savait presque par cœur.

Instance du motif : pour Mademoiselle de Baugé

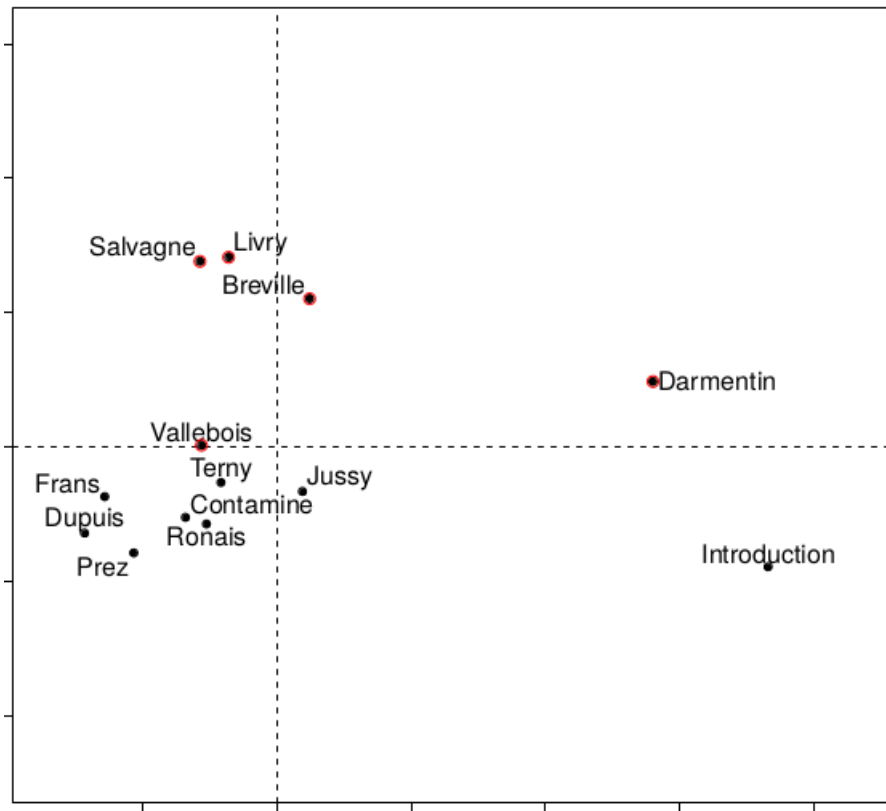
La phrase : Madame de Villiers voyant que c'était avec connaissance de cause que j'avais parlé surprise qu'il se défendît d'avoir **pour Mademoiselle de Baugé** un attachement, qui ne pouvait que c'était une très aimable personne, et qu'elle ne le croyait pas assez malhonnête homme pour

Enfin, ce motif des apocryphes confirme le fait que, dans ce conte, les personnages sont peu souvent appelés par leur nom propre.

Avec l'introduction ...

Sur la suggestion de Geneviève Artigas-Menant, le test est répété en ajoutant l'Introduction.
Vérifier la similarité entre l'introduction et l'Histoire de Madame Darmentin, qui en est une imitation.

(Attention, dans ce type d'analyse seulement la distribution relative est importante. Ici, l'algorithme décide, pour des raisons de lisibilité, de positionner les apocryphes en haut!)

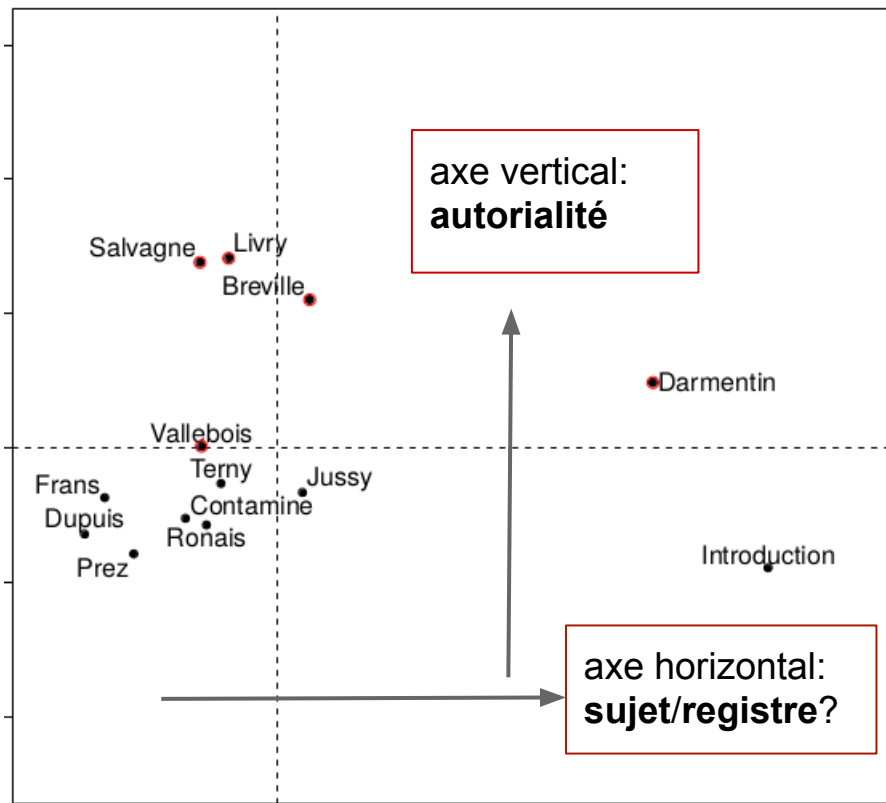


Avec l'introduction ...

Le test est répété en ajoutant l'Introduction.

Vérifier la similarité entre l'introduction et l'Histoire de Madame Darmentin, qui en est une imitation.

(Attention, dans ce type 'analyse seulement la distribution relative est importante. Ici l'algorithme décide, pour des raisons de lisibilité, de positionner les apocryphes en haut!)

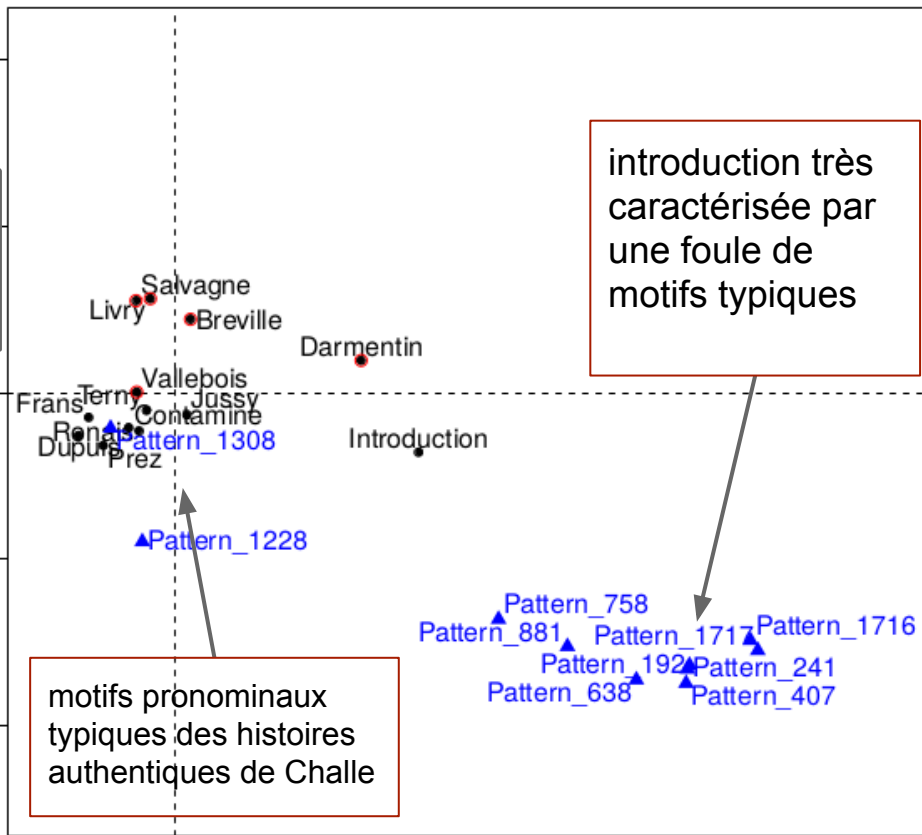


L'histoire de Mme D. et l'*Introduction* sont très proches et isolées des autres sur l'axe horizontal, ce qui semble donc correspondre à une similarité de sujet et/ou de registre.

L'axe vertical au contraire semble bien correspondre aux différences dans le style des auteurs.

Motifs de l'introduction ...

Visualisation des 10 motifs les plus importants.



Les motifs typiques de l'introduction sont tous liés à la séquence

PUN - VER - NAM - NAM - PUN

Motifs typiques de l'introduction

Voici quelques passages dans *L'introduction des Illustres Françaises* où l'on trouve la séquence [PUN - VER - NAM - NAM - PUN]

Instance du motif : , reprit Des Frans :

La phrase : Je ne refuse point vos offres , **reprit Des Frans** : ce qui m' y aurait pu obliger eût été la crainte de vous incommoder ; mais puisque vous m' assurez qu' il n' en sera rien , je reprends volontiers les anciens errements de notre amitié , et j' agirai avec vous sans façon .

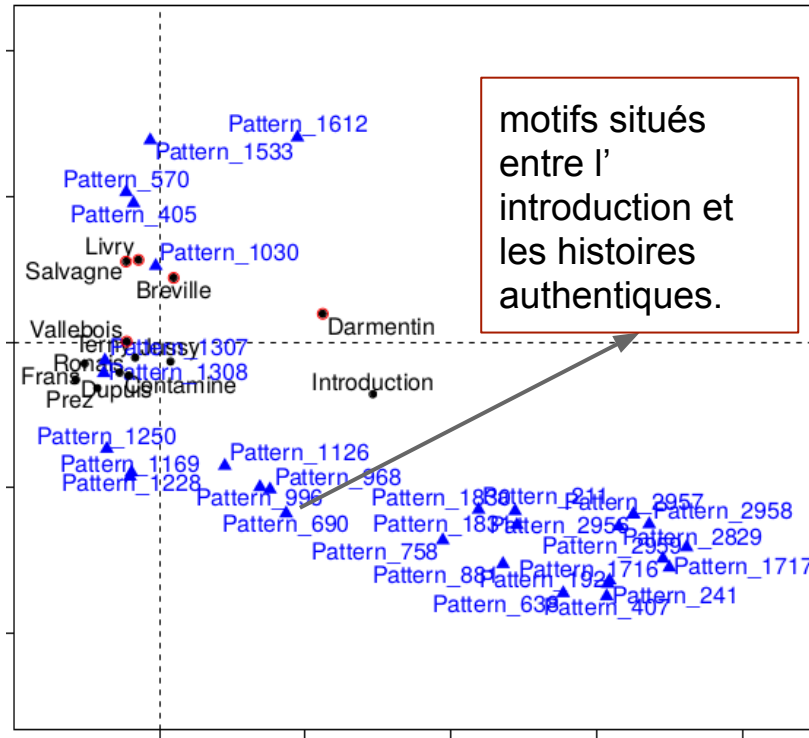
Instance du motif : , reprit Des Ronais ,

La phrase : C' est m' obliger , **reprit Des Ronais** , et vous ne me feriez pas plaisir d' en user autrement .

Instance du motif : , ajouta Des Frans ,

La phrase : Cela étant , **ajouta Des Frans** , ne trouvez pas mauvais que je ne sois point aujourd' hui des vôtres à dîner , ni peut-être encore à souper .

Motifs typiques de Challe



Pour récupérer les motifs chaliens on augmente le nombre des motifs imprimés: voici les **30 plus importants**.

Les motifs typiques de textes authentiques comme

Pattern_690:
PRO-PRO-VER-PUN-VER

Motifs typiques de Challe

Voici quelques passages dans *L'introduction des Illustres françaises* où l'on trouve la séquence [PRO-PRO-VER-PUN-VER]

Instance du motif : je l' entends , reprit

La phrase : C' est ainsi que **je l' entends , reprit** Des Ronais .

Instance du motif : où je suis , crotté

La phrase : Vous être le maître , dit Des Ronais , mais tout au moins , en attendant votre tailleur , vous boirez bien un coup à ma santé ; quatre si vous voulez , reprit Des Frans en riant , mais laissez -moi m' habiller ; car dans l' état **où je suis , crotté** et vilain , je me fais peur à moi-même .

Instance du motif : il vous plaira , répondit

La phrase : Vous le saurez quand **il vous plaira , répondit** le conseiller .

Instance du motif : Vous me surprenez , dit

La phrase : **Vous me surprenez , dit** Des Frans , de l' accuser d' infidélité , elle dont on vantait tant autrefois la sincérité et la candeur .

Motifs typiques de Challe

Voici quelques passages dans *L'introduction des Illustres françaises* où l'on trouve la séquence [PRO-PRO-VER-PUN-VER]

Instance du motif : je l' entends , reprit

La phrase : C' est ainsi que **je l' entends , reprit** Des Ronais .

Instance du motif : où je suis , crotté

La phrase : Vous être le maître , dit Des Ronais , mais tout au moins , en attendant votre tailleur , v
si vous voulez , reprit Des Frans en riant , mais laissez -moi m' habiller ; car dans l' état **où je suis**
même .

Instance du motif : il vous plaira , répondit

La phrase : Vous le saurez quand **il vous plaira , répondit** le conseiller .

Instance du motif : Vous me surprenez , dit

La phrase : **Vous me surprenez , dit** Des Frans , de l' accuser d' infidélité , elle dont on vantait tant autrefois la sincérité et la candeur .

Sur-utilisation de phrases
imbriquées, surtout mais pas
exclusivement dans le
dialogue.

Mais aussi structures
pronominales (narration à la
première personne!)

Conclusions préliminaires

Les textes authentiques semblent plus **homogènes**.

Ils se différencient des apocryphes par des traits communs, comme la préférence pour les motifs pronominaux (indice de la narration à la première personne) et des citations imbriquées.

Les apocryphes sont plus **éloignés** les uns des autres mais ils semblent se distinguer par certains traits qui sont moins présents dans les authentiques, comme la **répétition des noms propres par rapport aux locutions pronominales**, et le manque de noms propres.

Finalement, l'introduction aux Illustres Françaises **se détache des contes individuels** - tout en gardant des traits "challiens" - à cause de la répétition des noms propres et de la similarité avec le conte apocryphe de Madame Darmentin.