

Towards a computational and pattern-based stylistics



Jean-Gabriel Ganascia

ACASA team – LIP6

University Pierre et Marie Curie

OBVIL Labex

Sorbonne University

4, place Jussieu, 75252 Paris, Cedex 05

Jean-Gabriel.Ganascia@lip6.fr

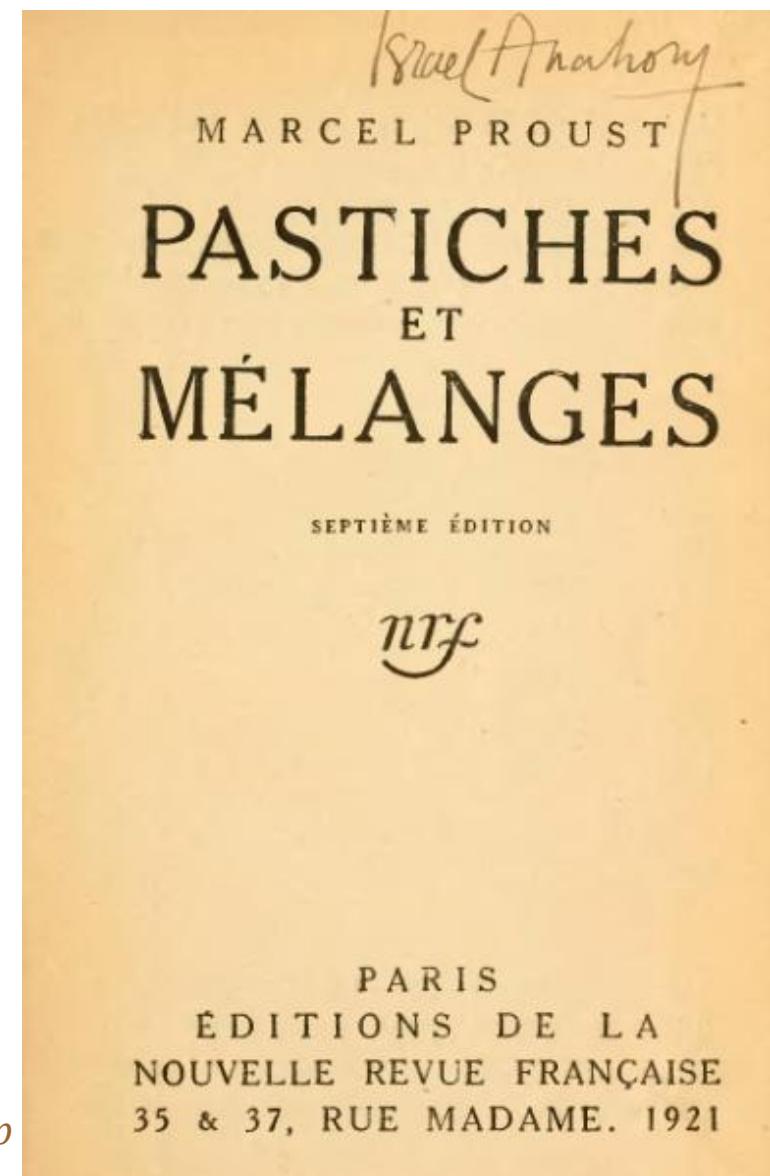


Jean-Gabriel GANASCIA



Synopsis

- What does mean “style”?
- Recurrent patterns in music: FlexPat
- Possible features of literary style
- The “Littératron”
- Syntactical patterns: EReMoS
- Semantical patterns



Style - Stylistic

- Etymology: stilus
- Characteristics
 - Genre (letter, theater, novel, ...)
 - Author
 - Characters (in drama)
 - Epochs
 - Gender
 -
- Conscious vs. unconscious?



Automatic Extraction of Patterns

UPMC

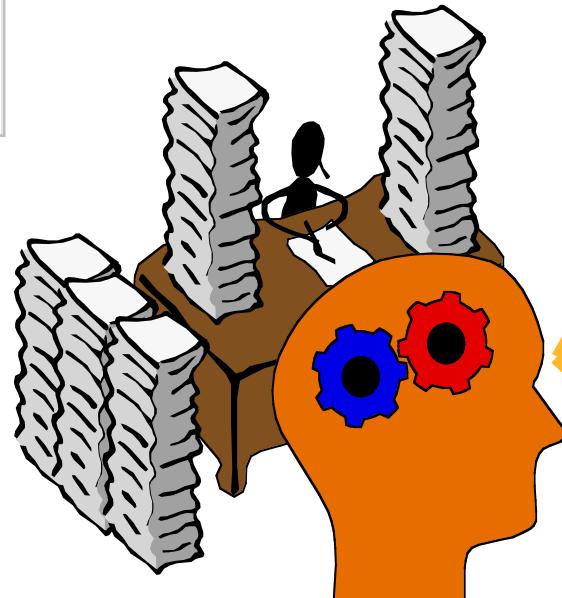
The image displays three horizontal staves of musical notation, likely from a score. Each staff is numbered at the top. The first staff has numbers 9, 10, 11, 12, 13, and 14. The second staff has numbers 22, 23, 24, and 25. The third staff has numbers 35, 36, 37, and 38. In each staff, a specific melodic pattern is circled with a thick orange oval. The first oval covers measures 10 through 13. The second oval covers measures 23 through 25. The third oval covers measures 35 through 38. The patterns consist of eighth and sixteenth note combinations, often involving grace notes or slurs.

Charlie
Parker
Solos



Lexicon of 200
melodic patterns

T. Owens (1974)

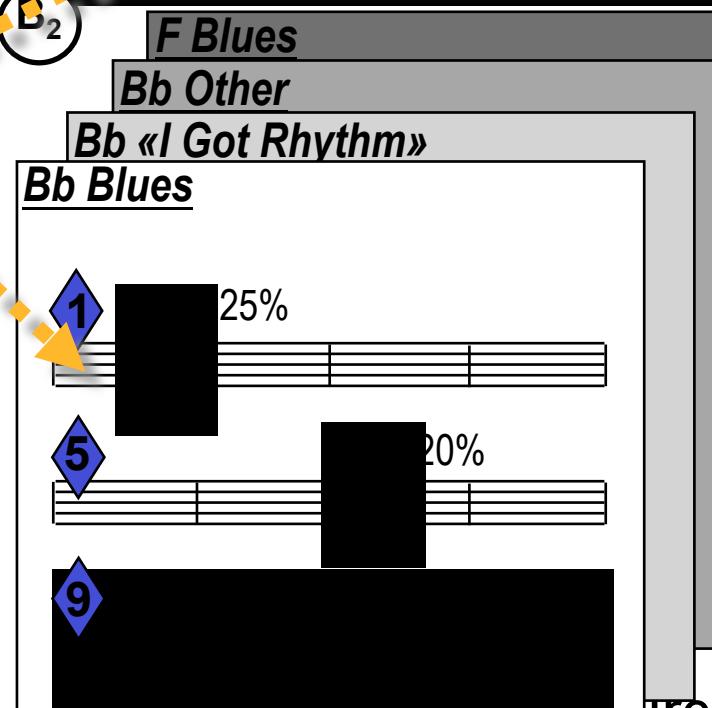


A

B₁

PATTERNS

	Global	Bb	F	C	...
1400	340	280	190	...	
1100	460	210	220	...	
850	140	230	60	...	



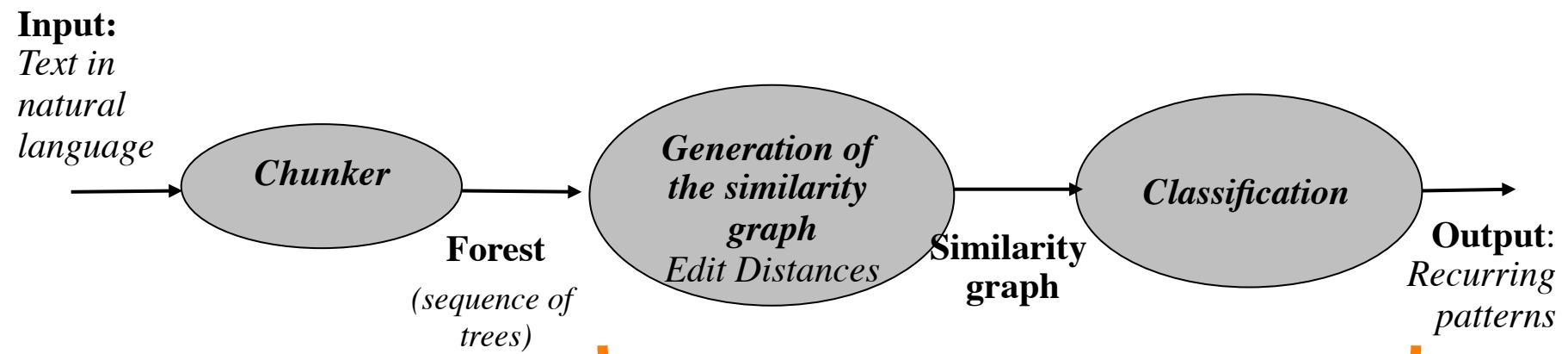
PhD Thesis P-Y Rolland (1998)
Multi-Valued Edit Model, i.e. linear
combination of Edit Distances

Features of Styles in Literary Studies

- Philology:
 - characteristics of an author: its syntax
- Lexicon
 - stop words → syntax
 - “heavy” words → semantics
- Syntaxical characteristics
- Rhythm (e.g. dactyl, iamb, ...) and punctuations
- Semantical characteristics: figures

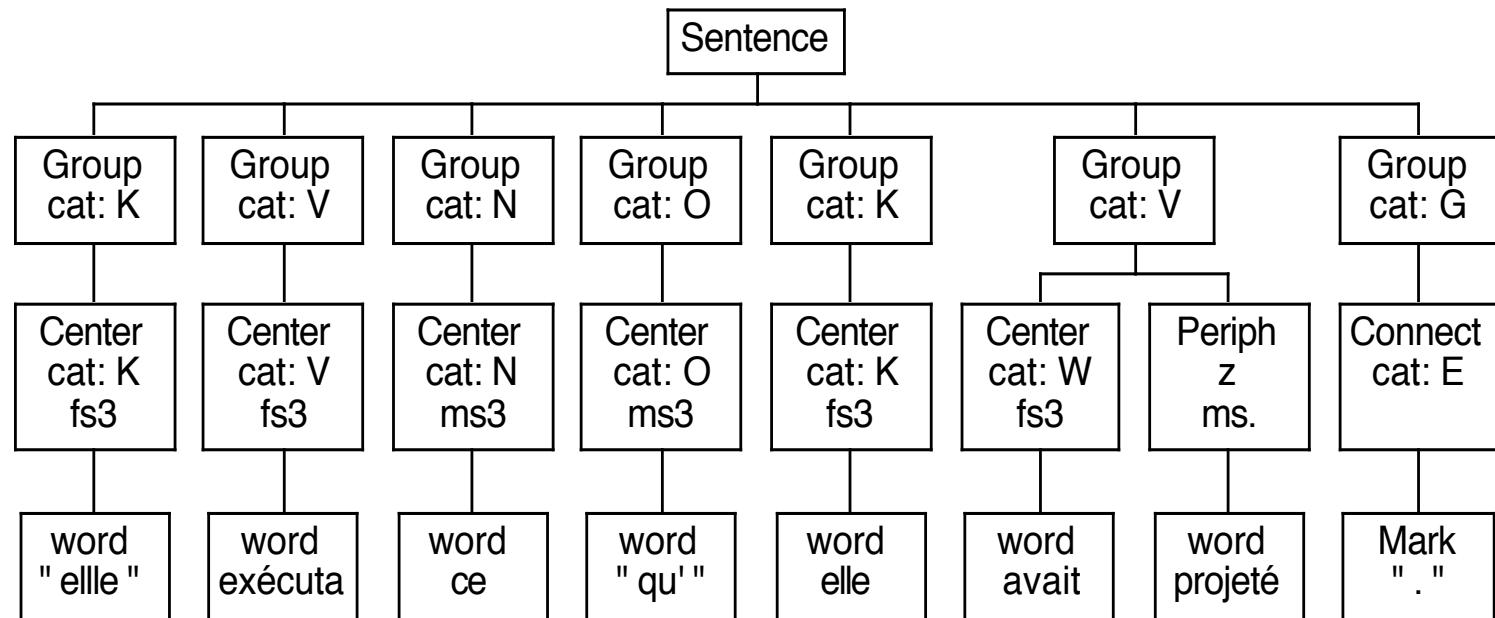


The “Littératron”*



Chunker

J. Vergnes



Elle exécuta ce qu'elle avait projeté.

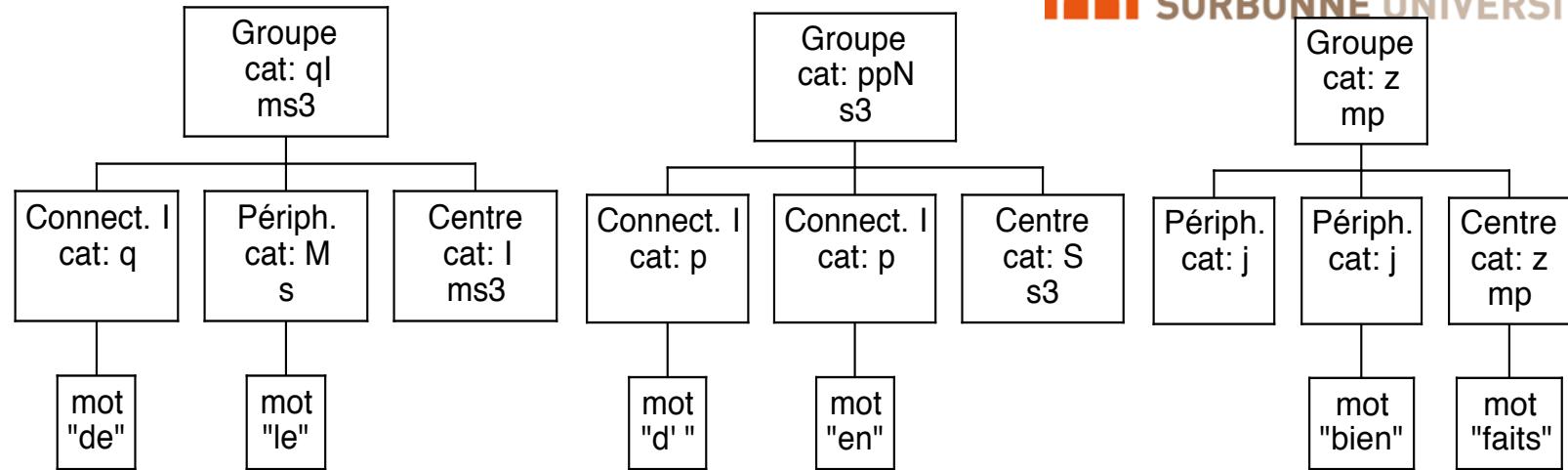
Madame de Lafayette (*La comtesse de Tende*)

Discrimination



- Patterns in Madame de Lafayette in the two following texts: “*La comtesse de Tende*” and “*La princesse de Clèves*”.
- And absent from many 19th century French authors : Guy de Maupassant, Georges Sand et Marcel Schwob.
Guy de Maupassant, *La peur* (1882), *La peur* (1884), *La veillée*, *La rempailleuse*, *Pierrot*, *En mer*, *Un normand*, *Ce cochon de Morin* and *Les sabots*, from Georges Sand, *La fée poussière*, *Le gnome des huîtres*, *Le marteau rouge*, *L'orgue du titan* and *La fée aux gros yeux*, from Marcel Schwob, *Arachné*, *Béatrice*, *Sur les dents*, *Le Dom*, *L'homme double*, *Le fort*, *Gabelous*, *Parabole*, *Lilith*, *Conte des œufs* and *Les portes de l'opium*.

Three specific patterns

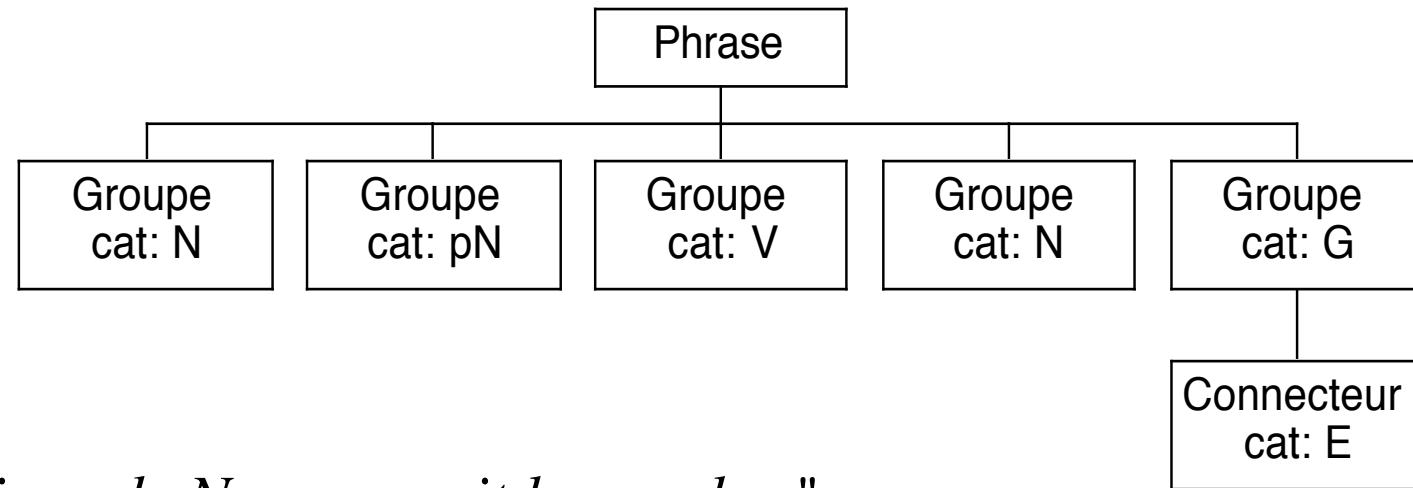


The **first** covers the following expressions : "*à le servir*", "*de le supplier*", "*de l'éviter*", "*de l'aimer*"...

The **second**, "*sans en avoir*", "*d'en avoir*", "*d'en attendre*", "*d'en garantir*", "*d'en faire*"...

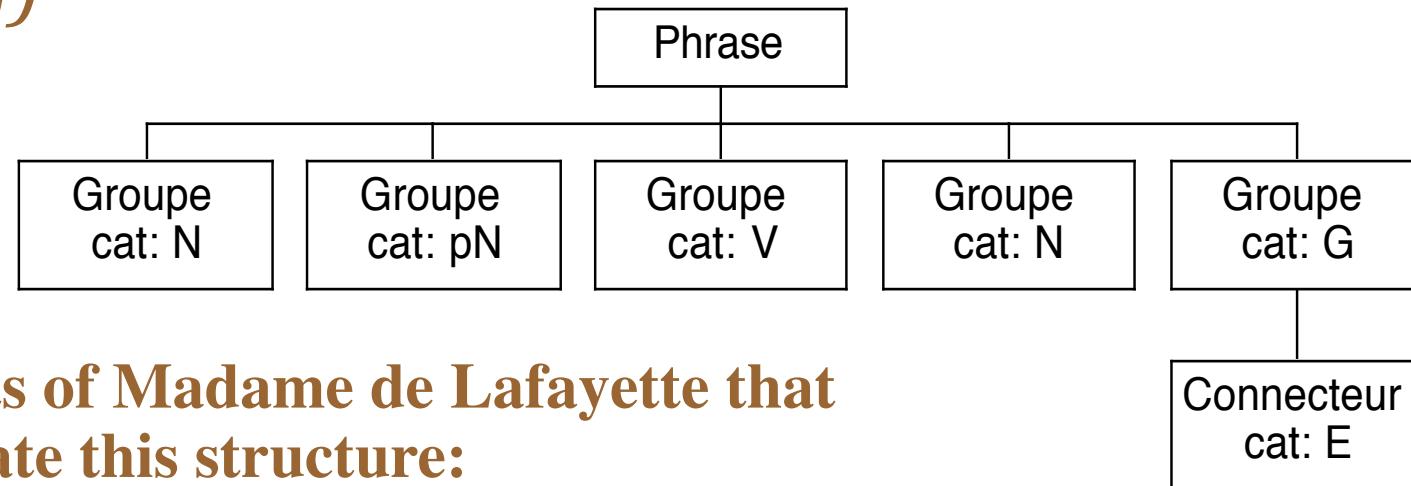
The **third** covers three fragments : "*admirablement bien faits*", "*parfaitement bien faits*", "*très bien fait*"

Typical Structure of Madame de Lafayette



"*Le prince de Navarre prit la parole :*",
"*La reine de Navarre avait ses favorites*",
"*Madame de Clèves ne répondit rien*",
"*Monsieur de Nemours prit la reine dauphine*",...

Typical Structure of Madame de Lafayette (following)



Fragments of Madame de Lafayette that approximate this structure:

"*Madame de Chartres avait une opinion opposée ;*",

"*Le comte de Tende aimait déjà le chevalier de Navarre ;*",

"*Le comte de Tende sentit son procédé dans toute sa dureté ;*",

"*La comtesse reçut ce billet avec joie*", ...

EReMoS

Acronym for “*Extraction et Recherche de Motifs Syntaxiques*”



- Lemmatization of text
- Extraction of frequent sequences of lemmas (i.e. POS tags)
- <http://eremos.lip6.fr/>
- Available for French and German and soon for Spanish and English
- “Le silence profond régnait nuit et jour dans la maison.”
 - <DET, NOM, ADJ, VER, NOM, KON, NOM, PRP, DET, NOM, SENT>
- Sequential patterns:
 - <DET><NOM><ADJ>
 - <NOM><ADJ><VER><NOM>
 - <KON><NOM><*><DET><NOM>

Dictionnaire Le Robert | Le Grand Robert de la langue fr... | EReMoS - Accueil

eremos.lip6.fr | Rechercher | +

EREMOS (FR) | EREMOs -(GERMAN) (DE)

AIDE + | À PROPOS

Observatoire
de la vie littéraire

Fréquence d'apparition relative des motifs: 10 %

Fréquence d'apparition absolue des motifs: 10

Longueur des motifs:

minimale: 3

maximale: 5

Aucun fichier | PARCOURIR

Veuillez charger un texte (seul les fichiers txt en encodage UTF-8 sont acceptés), le nom du fichier ne doit pas contenir des caractères spéciaux ou d'espaces.

ANNULER | ENVOYER

eremos.lip6.fr

Applications Bookmarks Dictionnaires Google Agenda Courriel LIP6 Crédit Mutuel Cybernity Visualisation DH Vélos Centre de Production Autres favoris

EREMOS (FR) EREMO - (GERMAN) AIDE + À PROPOS Observatoire de la vie littéraire

= I. Salon de 1845 =

== I. Quelques mots d'introduction ==

Nous pouvons dire au moins avec autant de justesse qu'un écrivain bien connu à propos de ses petits livres : ce que nous disons, les journaux n'oseraient l'imprimer. Nous serons donc bien cruels et bien insolents ? non pas, au contraire, impartiaux. Nous n'avons pas d'amis, c'est un grand point, et pas d'ennemis. — Depuis M. G. Planche, un paysan du Danube dont l'éloquence impérative et savante s'est tue au grand regret des sains esprits, la critique des journaux, tantôt niaise, tantôt furieuse, jamais indépendante, a, par ses mensonges et ses camaraderies effrontées, dégoûté le bourgeois de ces utiles guide-ânes qu'on nomme comptes rendus de Salons[1].

Et tout d'abord, à propos de cette impertinente appellation, le ~~bourgeois~~, nous déclarons que nous ne partageons nullement les

baudelaire_curiosités-es PARCOURIR

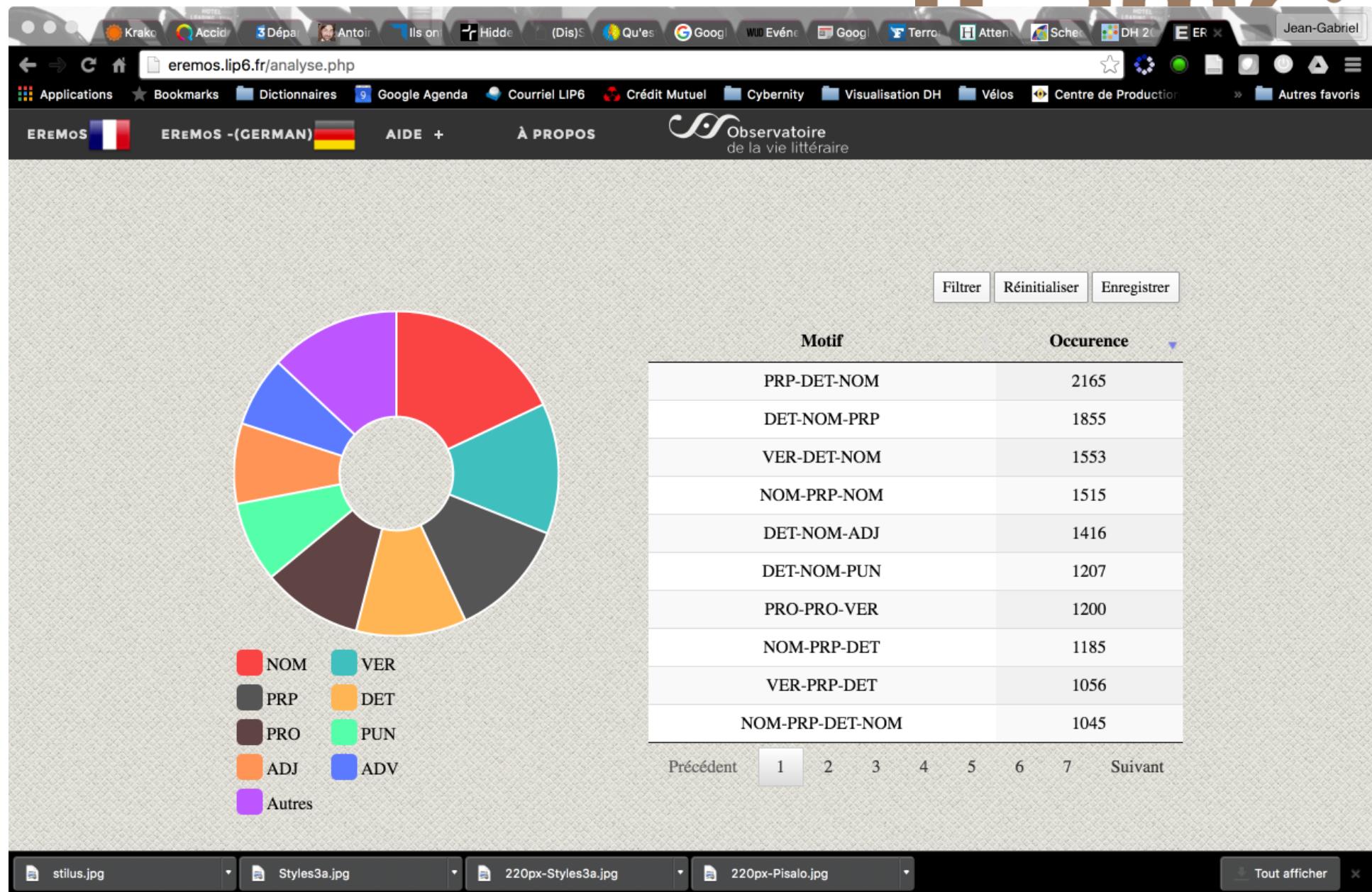
Veuillez charger un texte (seul les fichiers txt en encodage UTF-8 sont acceptés), le nom du fichier ne doit pas contenir des caractères spéciaux ou d'espaces.

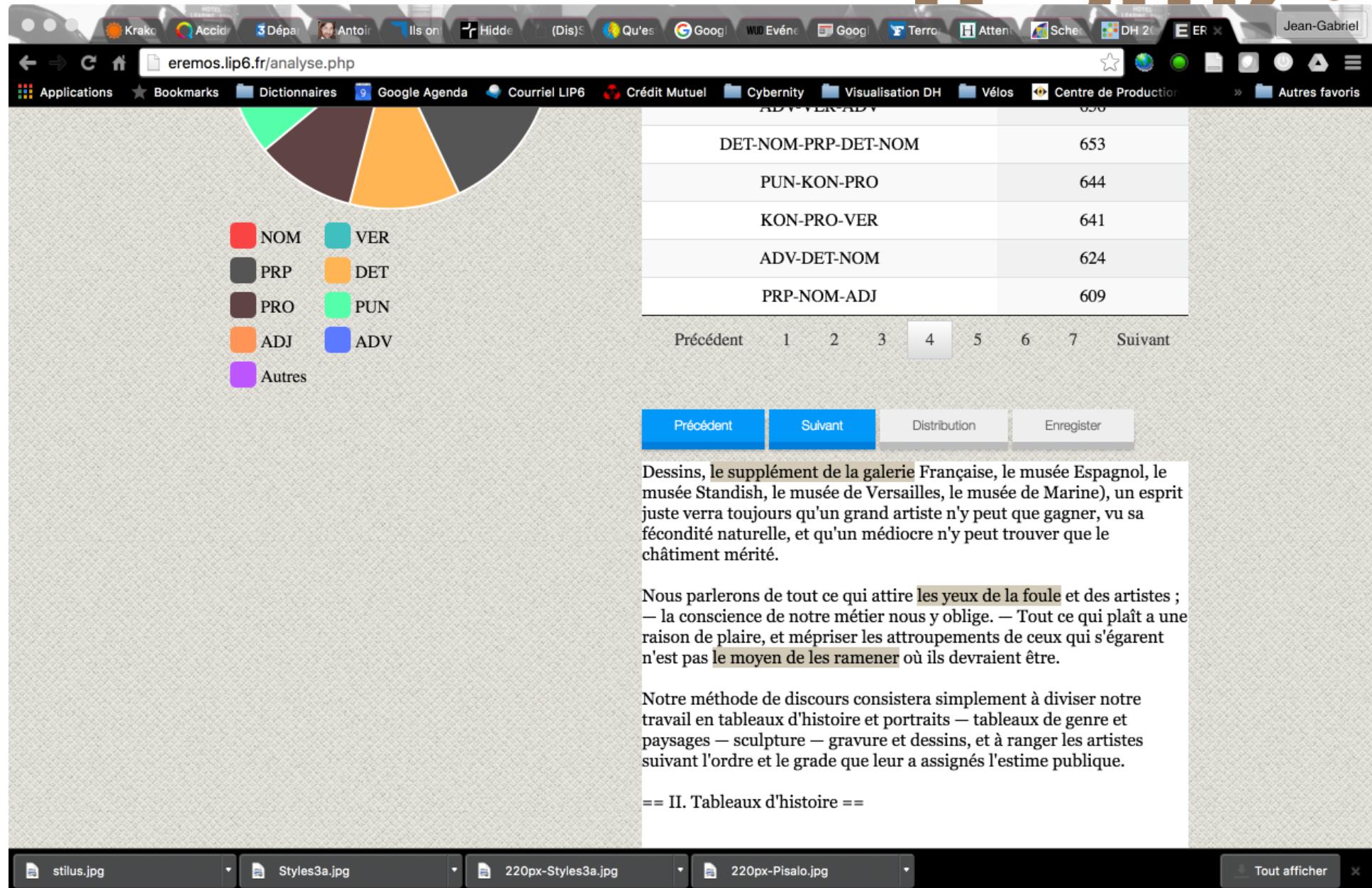
ANNULER ENVOYER

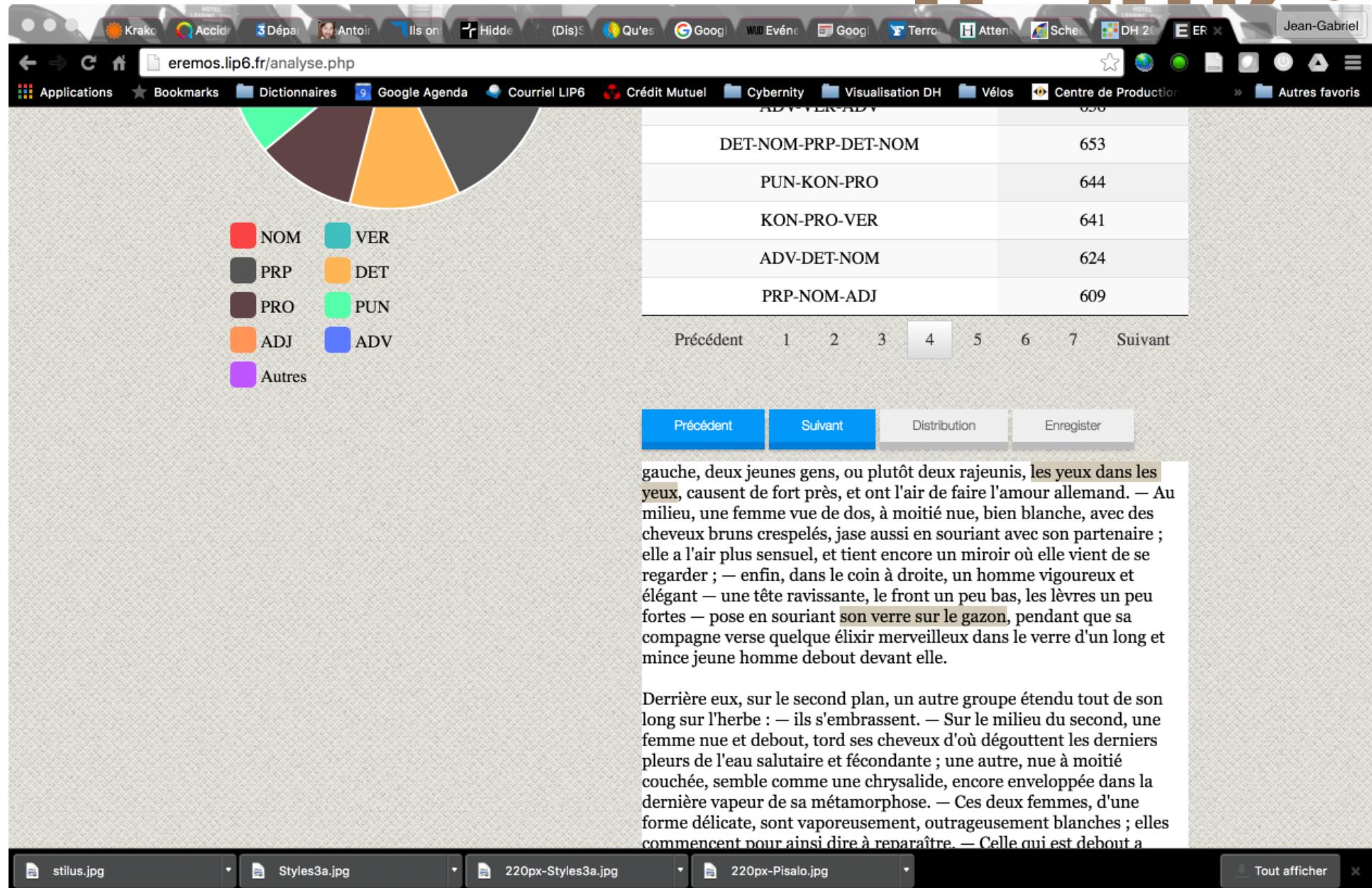
stilus.jpg Styles3a.jpg 220px-Styles3a.jpg 220px-Pisalo.jpg Tout afficher

Fréquence d'apparition relative des motifs	<input type="text" value="10"/> %	?
Fréquence d'apparition absolue des motifs	<input type="text" value="10"/>	
Longueur des motifs		
minimale	<input type="text" value="3"/>	
maximale	<input type="text" value="5"/>	



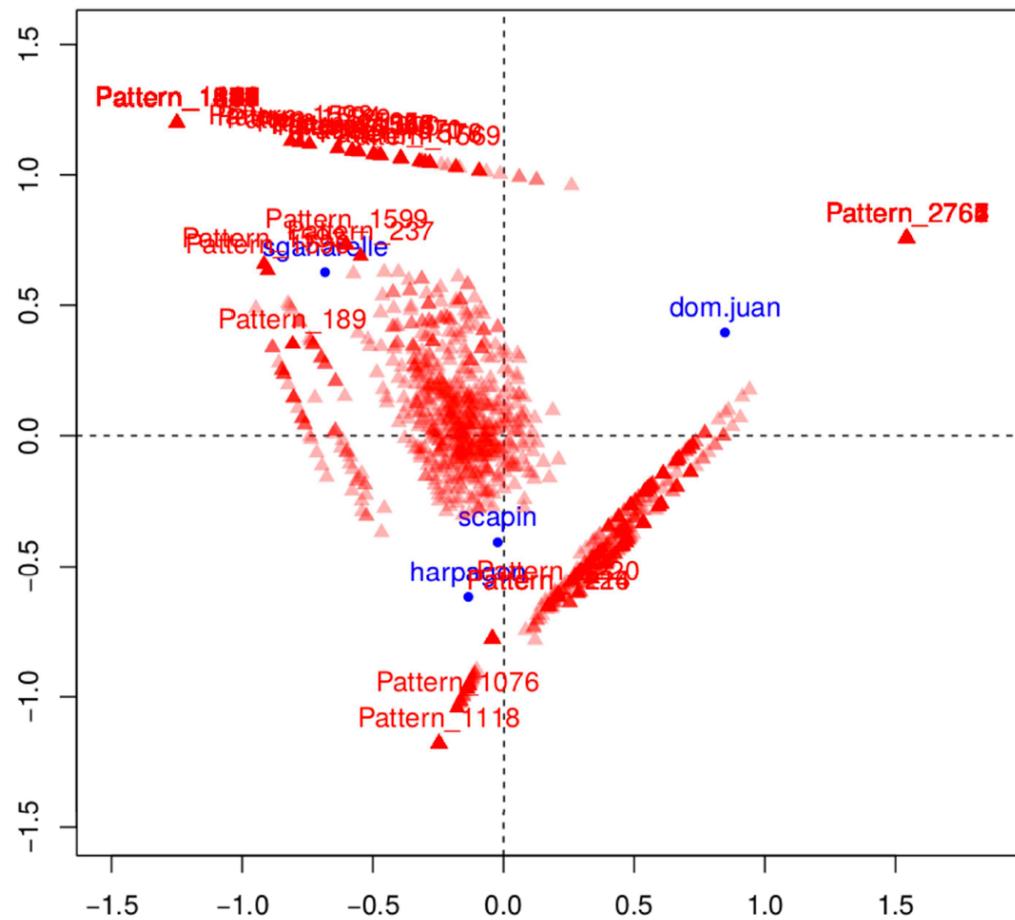






Memorable Molière's Protagonists

Boukhaled, Besnard, Frontini 2015



- Don Juan
- Sganarelle
- Scapin
- Harpagon

“interestingness” measure

Boukhaled 2016

- Distribution of patterns in the text
- $D = (d_1, d_2, \dots, d_{n-1})$
- d_i distance between two occurrences
- $\sigma_{norm} = \sigma / \sigma_{geo}$

$$\sigma = \frac{\sqrt{E(D^2) - E(D)^2}}{E(D)}$$

$$\sigma_{geo} = \sqrt{1 - \frac{1}{E(D)}}$$

<PUN><ADV><PRO> (119 occurrences)



<PUN><NOM><NAM> (120 occurrences)



Positions of the pattern in the text

Distribution of patterns



<ADV> <PRO> <NOM> (SUPPORT= 109)



<NAM> <PRP> <NAM> <PUN> (SUPPORT= 118)



<PRO> <VER> <PRO> <PRO> (SUPPORT= 128)



<NUM> <NOM> <PRP> <NOM> (SUPPORT= 117)



<DET> <NOM> <SENT> (SUPPORT= 2480)

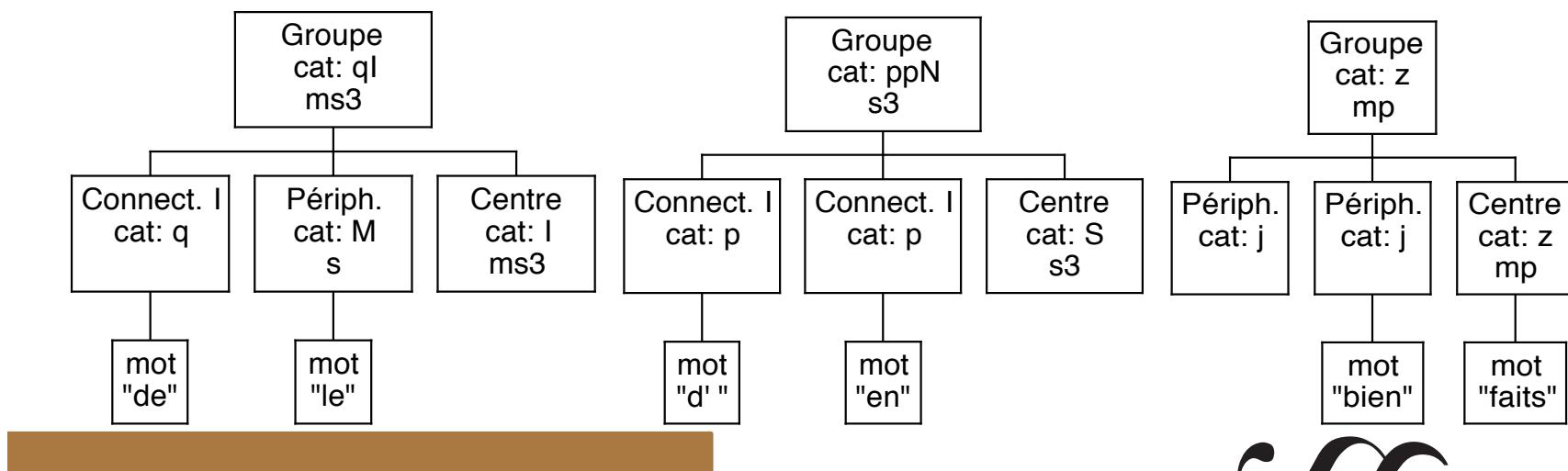


<NAM> <PRP> <NAM> <SUP> (SUPPORT= 215)



Perspective: extraction of “non flat” patterns

- Sequential flat patterns:
 - <DET><NOM><ADJ>
 - <NOM><ADJ><VER><NOM>
 - <KON><NOM><*><DET><NOM>
- Non flat patterns: mixing *lemmas, words, categories, variables*
 - <DET><“le”><NOM><ADJ>
 - <KON><NOM><*><DET><NOM>



Detection of Figures

e.g. Extraction of simile

Suzanne Mpouli



And then in the fall the crackling of cornleaves, the smells, sights and sounds.

The corn stood up like armies in the shocks.

When I was a boy I went into the cornfields at night. I said words I had not dared to say to people, defying the New Englanders' gods, trying to find honest, mid-western American gods.

Sherwood Anderson (1876-1941) – “Mid-American Prayer”.

