



Séries textuelles chronologiques

André Salem

Université de la Sorbonne nouvelle - Paris 3

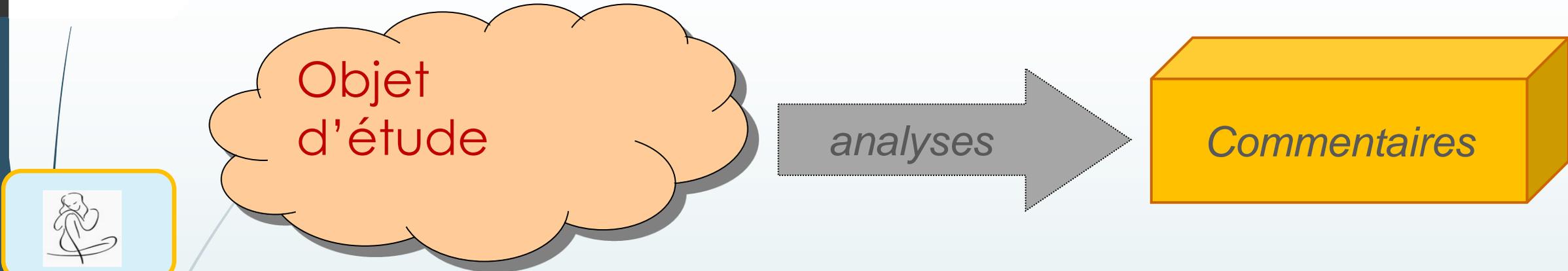
Plan

- introduction : *méthodes et illusions*
- textométrie - statistique textuelle
 - domaines d'application
 - méthodes
- séries textuelles chronologiques
- méthodes de la textométrie chronologique
- conclusions / perspectives



Méthodes et illusions

Méthodes et illusions 3

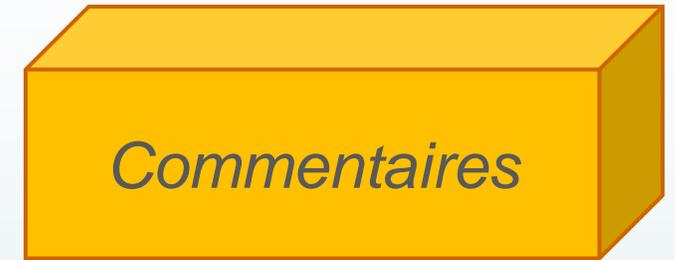
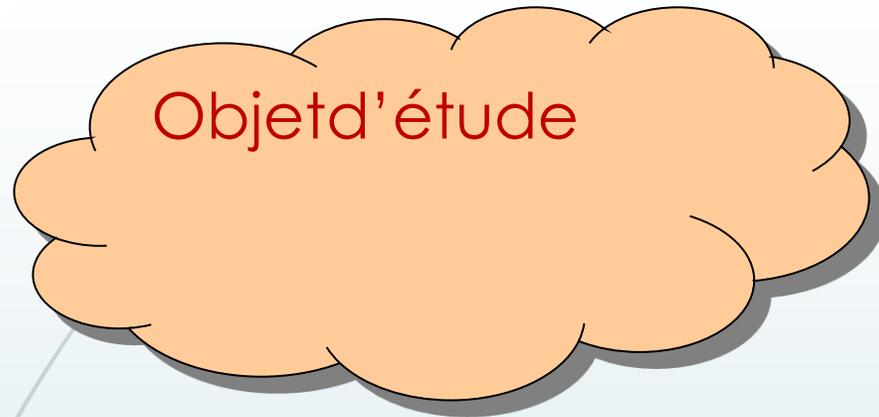


Objet
d'étude

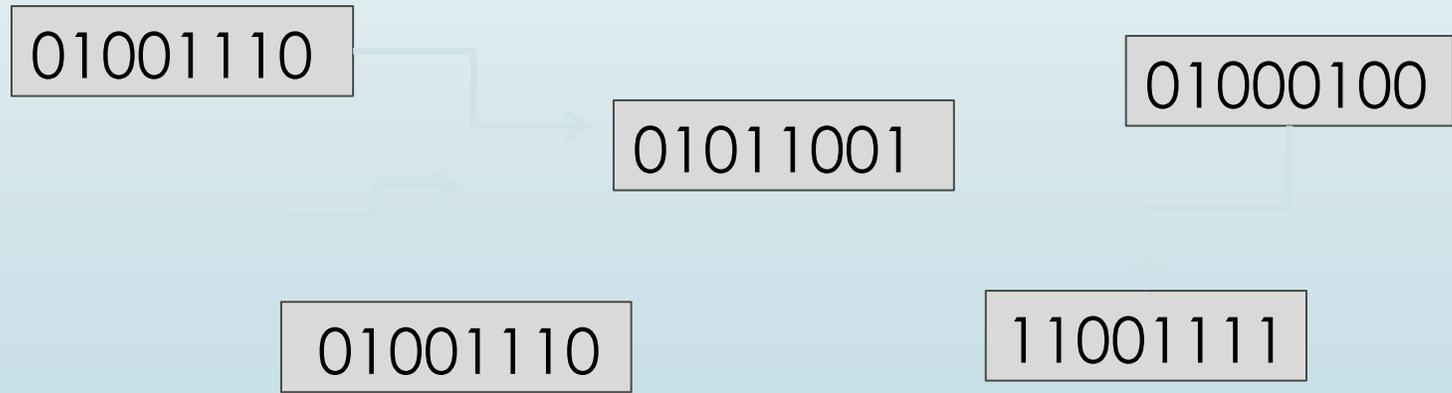
analyses

Commentaires

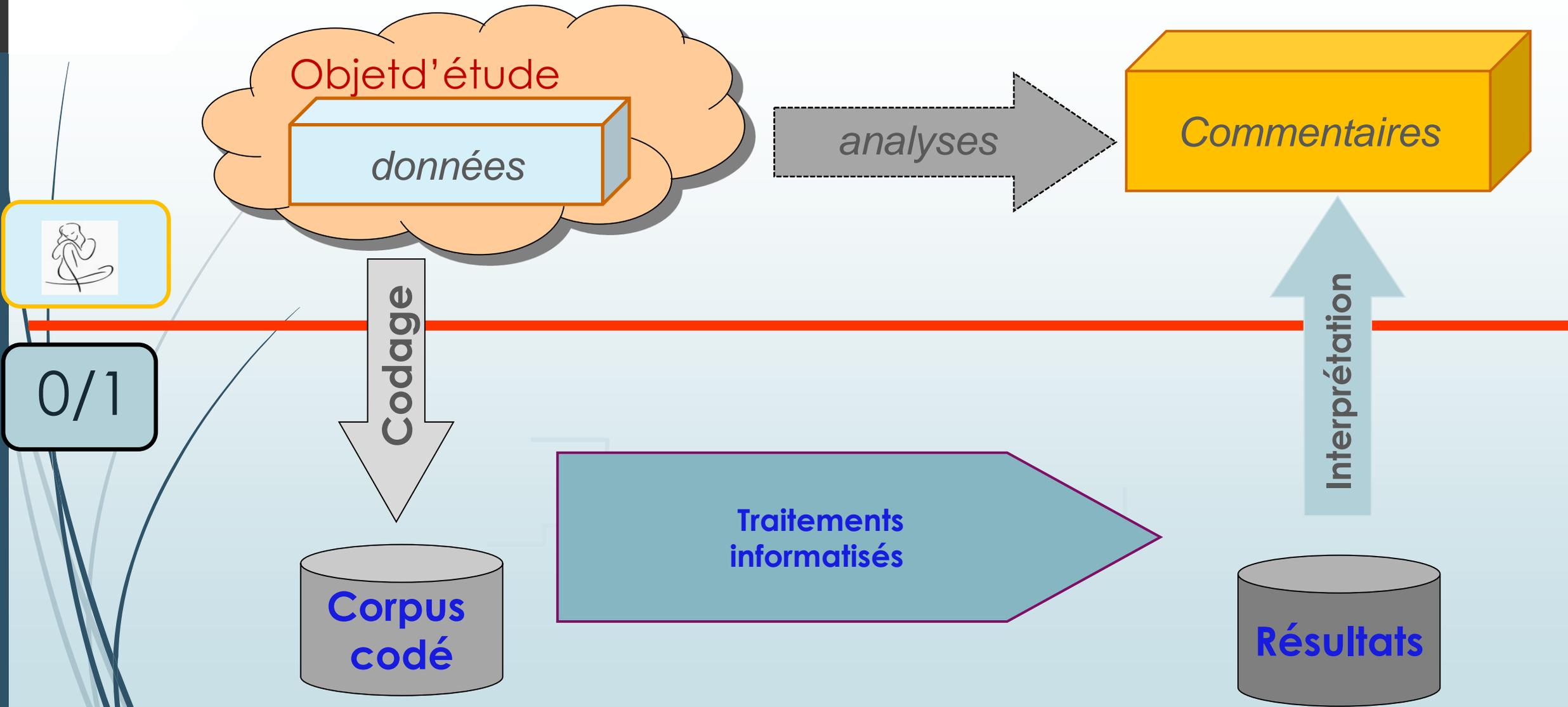
Méthodes et illusions 3



0/1



Méthodes et illusions 3



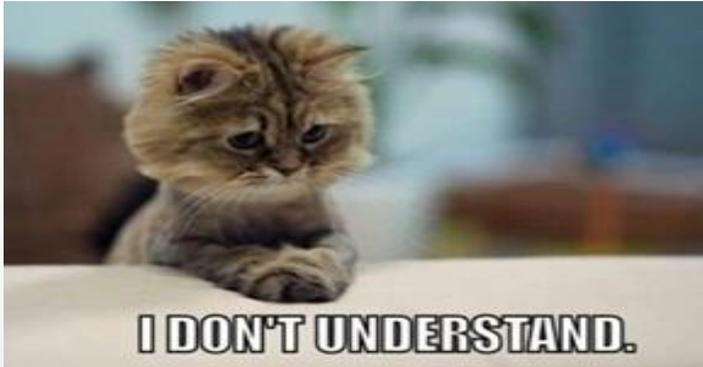
Méthodes et illusions 4



0110101 101010110010110101

0/1

Méthodes et illusions 4



Alice was beginning to get very tired of ...

011011001010011000000

0/1

0110101 101010110010110101

Méthodes et illusions 4



Alice was beginning to get very tired

0/1

Alice was beginning to get very tired

Alice was beginning to get very tired

Méthodes et illusions 4



Alice was beginning to get very tired



0/1

Alice was beginning to get very tired

Alice was beginning to get very tired

NP

VBD

VVG

TO

VV

RB

JJ



Domaines d'application

quantitative studies :

1937 Handley HL - *Word Index to James Joyce's « Ulysses »*

<Chapter=1>

§ Stately, plump Buck Mulligan came from the stairhead, bearing a bowl of lather on which a mirror and a razor lay crossed. A yellow dressinggown, ungirdled, was sustained gently behind him on the mild morning air. He held the bowl aloft and intoned:

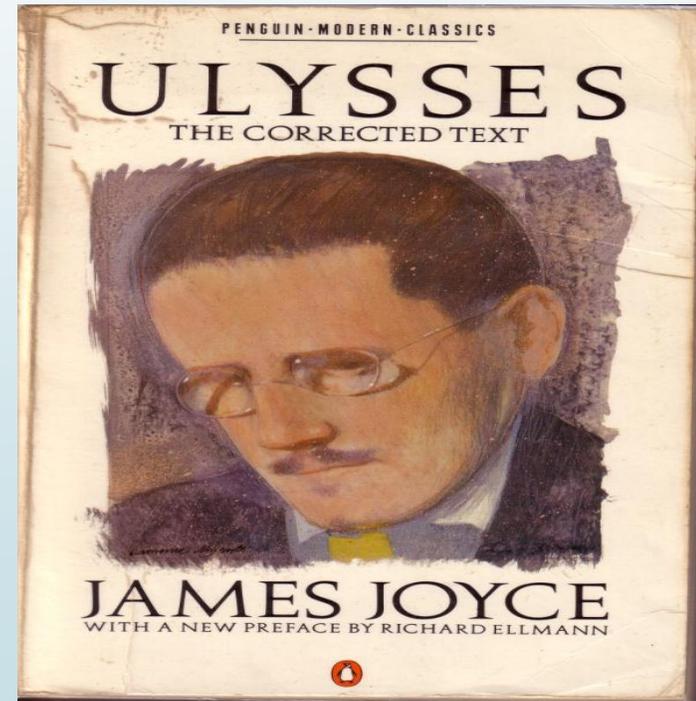
—*Introibo ad altare Dei.*

Halted, he peered down the dark winding stairs and called out coarsely:

—Come up, Kinch! Come up, you fearful jesuit!

Solemnly he came forward and mounted the round gunrest. He faced about and blessed gravely thrice the tower, the surrounding land and the awaking mountains. Then, catching sight of Stephen Dedalus, he bent towards him and made rapid crosses in the air, gurgling in his throat and shaking his head. Stephen Dedalus,

tokens : 260 000
types : 33 692



polémiques : *théâtre classique*

Corneille, Racine, Molière

1919-2010 Pierre Louÿs, Marcel Cohen, D. Labbé, G. Forestier,
... et beaucoup d'autres

Amphitryon

La scène est à Thèbes, devant la maison d'Amphitryon.

PROLOGUE

MERCURE, sur un nuage; LA NUIT, dans un char traîné par deux chevaux.

MERCURE

Tout beau, charmante Nuit; daignez vous arrêter.

Il est certain secours, que de vous on désire:

Et j'ai deux mots à vous dire,

De la part de Jupiter.

LA NUIT

Ah, ah, c'est vous, Seigneur Mercure!

Qui vous eût deviné là, dans cette posture?

L'École des femmes, I, 1

Je sais un paysan qu'on appelait Gros-Pierre

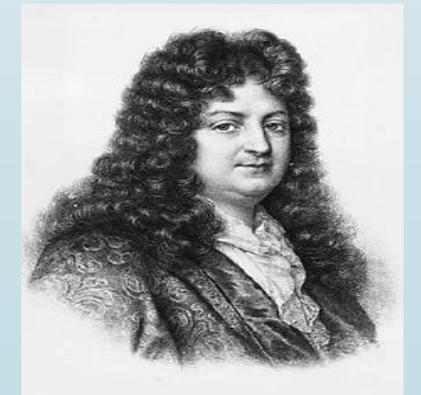
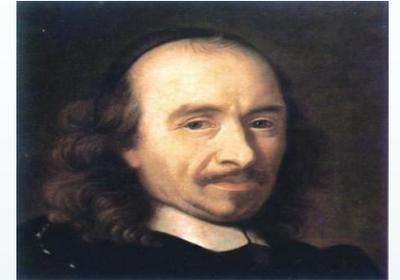
Qui n'ayant pour tout bien qu'un seul quartier de terre,

Y fit tout à l'entour faire un fossé bourbeux,

Et de Monsieur de l'Île en prit le nom pompeux.

Attribution

Richesse du vocabulaire



application areas : *historical studies*

1975 *Le Père Duchesne de Hébert* (1793-1794) (série textuelle chronologique)

tokens : 140 000
types : 11 000

<mois=01><numero=260> <Epg=1>

§ la grande douleur du *père *duchesne au sujet de la mort de *marat assassiné à coups de couteau par une garce du *calvados, dont l'évêque *fauchet était le directeur. ses bons avis aux braves *sans-culottes pour qu'ils se tiennent sur leurs gardes, attendu qu'il y a dans *paris plusieurs milliers de tondus de la *vendée qui ont la patte graissée pour égorger tous les bons citoyens.

§ *marat n'est plus, foutre. peuple, gémis, <Epg=2> pleure ton meilleur ami; il meurt martyr de la liberté. c'est le *calvados qui a vomi le monstre sous les coups duquel il vient de périr. une jeune fille, ou plutôt une furie armée par les prêtres, et pénitente, dit on, du cafard *fauchet, part de *caen pour exécuter cet horrible attentat. elle arrive à *paris, et après avoir acheté un large couteau au



Je suis le véritable Père Duchesne, foutre !

LA GRANDE JOIE DU PÈRE DUCHESNE.

APRES avoir vu défiler la procession des Brissotins, des Girondins et des Rolandins, pour aller jouer à la main chaude à la place de la Révolution. Le testament de Cartouche Brissot, et la confession du prêtre Fauchet qui a fait le caffard jusqu'à la fin, pour faire pleurer les vieilles dévotes, mais qui, dans le fond du cœur, se foutoit autant du père éternel que du grand diable Belzébuc.

ADIEU paniers, vendanges sont faites; tous

application areas : *political studies*

2015 Souza S - *Fidel Castro's speeches (1959-2008)*

<texte=0001> <date=1959_01_01> §

santiagueros;§ compatriotas de toda cuba:§

al fin hemos llegado a santiago. duro y largo ha sido el camino, pero hemos llegado.§

se decía que hoy a las 2 de la tarde se nos esperaba en la capital de la república, el primer extrañado fui yo, porque yo fui uno de los primeros sorprendidos con ese golpe traidor y amañado de esta mañana en la capital de la república.§

además, yo iba a estar en la capital de la república, o sea, en la nueva capital de la república, porque santiago de cuba será, de acuerdo con el deseo del presidente provisional, de acuerdo con el deseo del ejército rebelde y de acuerdo con el deseo del pueblo de santiago de cuba, que bien se lo merece, la capital. ¡santiago de cuba será la capital provisional de la república!§

LHT - Paris janvier 2019

tokens : 8 000 000
types : 77 000



application areas : *political studies*

2017

Miao J - Salem A

Congress of Communist Party of China (1982-2017)

<cong=1982>

#全面开创社会主义现代化建设的新局面
#--胡耀邦在中国共产党第十二次全国代表大会上的报告 #（一九八二年九月一日）

\$同志们：

#现在，我代表中国共产党第十一届中央委员会，向党的第十二次全国代表大会作报告。

\$一*历史性的转变和新的伟大任务

#自从一九七六年十月粉碎江青反革命集团以来，特别是党的十一届三中全会以来，经过全党全军全国各族人民的艰苦努力，我们已经在指导思想上完成了拨乱反正的艰巨任务，在各条战线的实际工作中取得了拨乱反正的重大胜利，实现了历史性的伟大转变。

tokens : 115 138
types : 7 365



application areas : *philological studies*

2015 The Kuran

<Sourate=001>
<v=1> # بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ
<v=2> # اَلْحَمْدُ لِلّٰهِ رَبِّ الْعٰلَمِیْنَ
<v=3> # الرَّحْمٰنِ الرَّحِیْمِ
<v=4> # مَا لِكَ یَوْمَ الدِّیْنِ
<v=5> # اِیَّاكَ نَعْبُدُ وَاِیَّاكَ نَسْتَعِیْنُ
<v=6> # اِهْدِنَا الصِّرَاطَ الْمُسْتَقِیْمَ
<v=7> # صِرَاطَ الَّذِیْنَ اَنْعَمْتَ عَلَیْهِمْ غَیْرِ الْمَغْضُوْبِ عَلَیْهِمْ وَلَا الضَّالِّیْنَ
<Sourate=002>
بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ
<v=1> # اَلْم
<v=2> # ذٰلِكَ الْكِتٰبُ لَا رَیْبَ فِیْهِ هُدًى لِّلْمُتَّقِیْنَ
<v=3> # الَّذِیْنَ یُؤْمِنُوْنَ بِالْغَیْبِ وَیُقِیْمُوْنَ الصَّلٰةَ وَمِمَّا رَزَقْنٰهُمْ یُنْفِقُوْنَ
وَالَّذِیْنَ یُؤْمِنُوْنَ بِمَا اُنزِلَ اِلَیْكَ وَمَا اُنزِلَ مِنْ قَبْلِكَ وَبِالْآخِرَةِ هُمْ یُوقِنُوْنَ

types : 77 874
tokens : 17 503

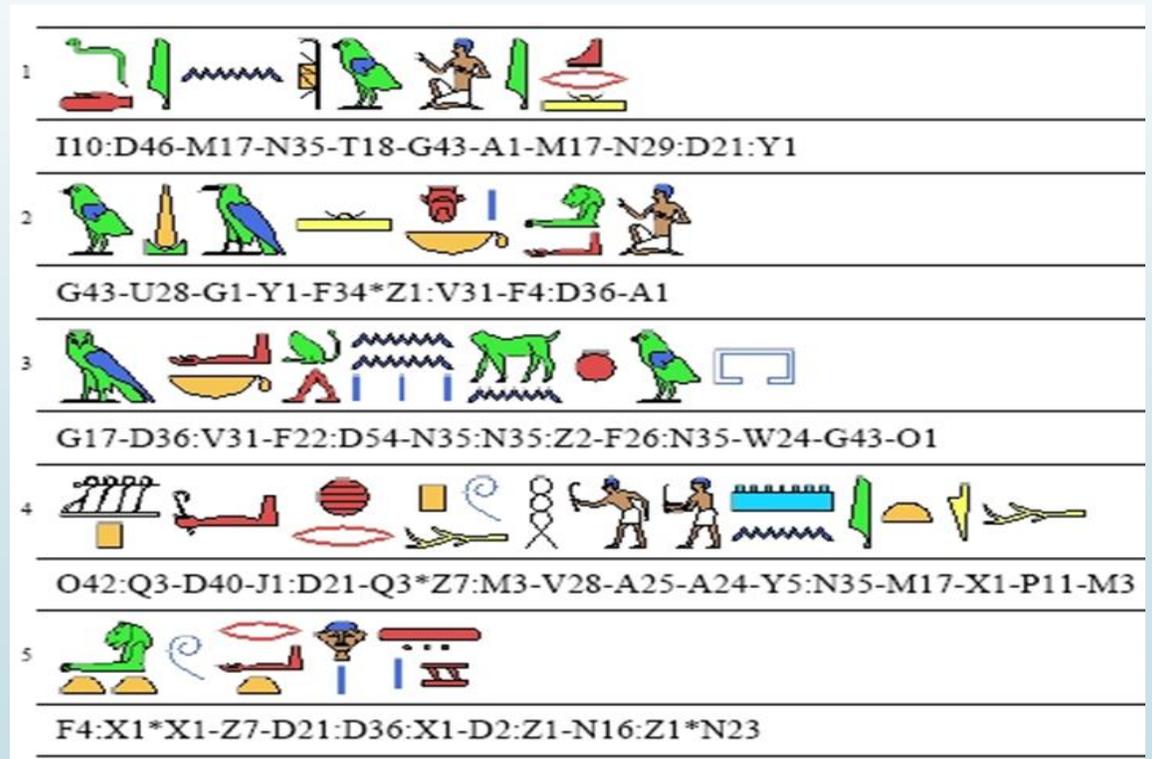


application areas : *literary studies*

2008 The Shipwrecked Sailor (2200 BCE)

THE wise servant said, "Let thy heart be satisfied, O my lord, for that we have come back to the country; after we have been long on board, and rowed much, the prow has at last touched land. All the people rejoice and embrace us one after another. Moreover, we have come back in good health, and not a man is lacking; although we have been to the ends of Wawat [Nubia], and gone through the land of Senmut [Kush], we have returned in peace, and our land---behold, we have come back to it. Hear me, my lord; I have no other refuge. Wash thee, and turn the water over thy fingers; then go and tell the tale to the majesty."

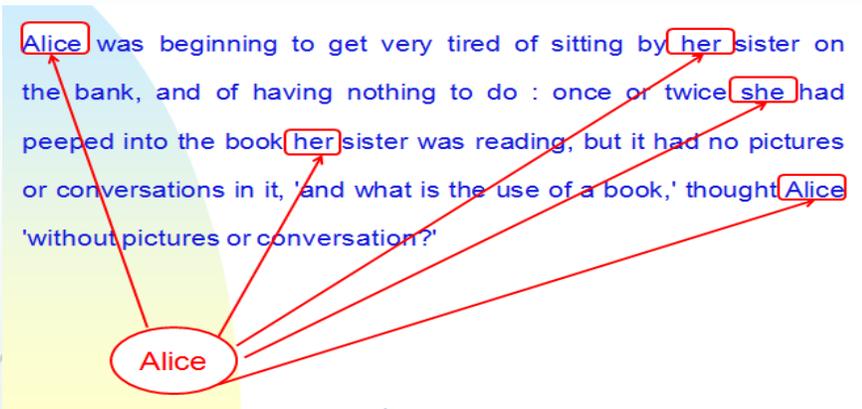
tokens : 3 700
types : 248





Méthodes

Méthodes documentaires ...



Localisation

Méthodes statistiques ...

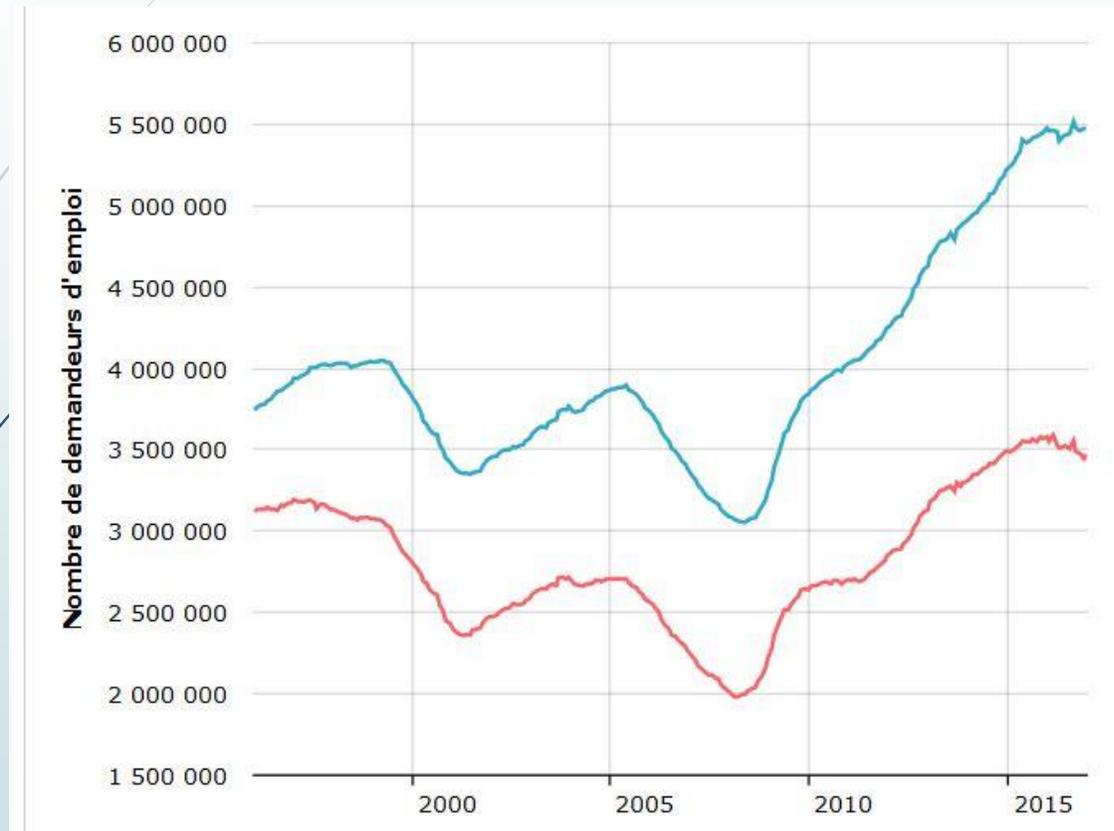
Figures don't lie, but liars do figure.

M. Twain

There are lies, damned lies and statistics.

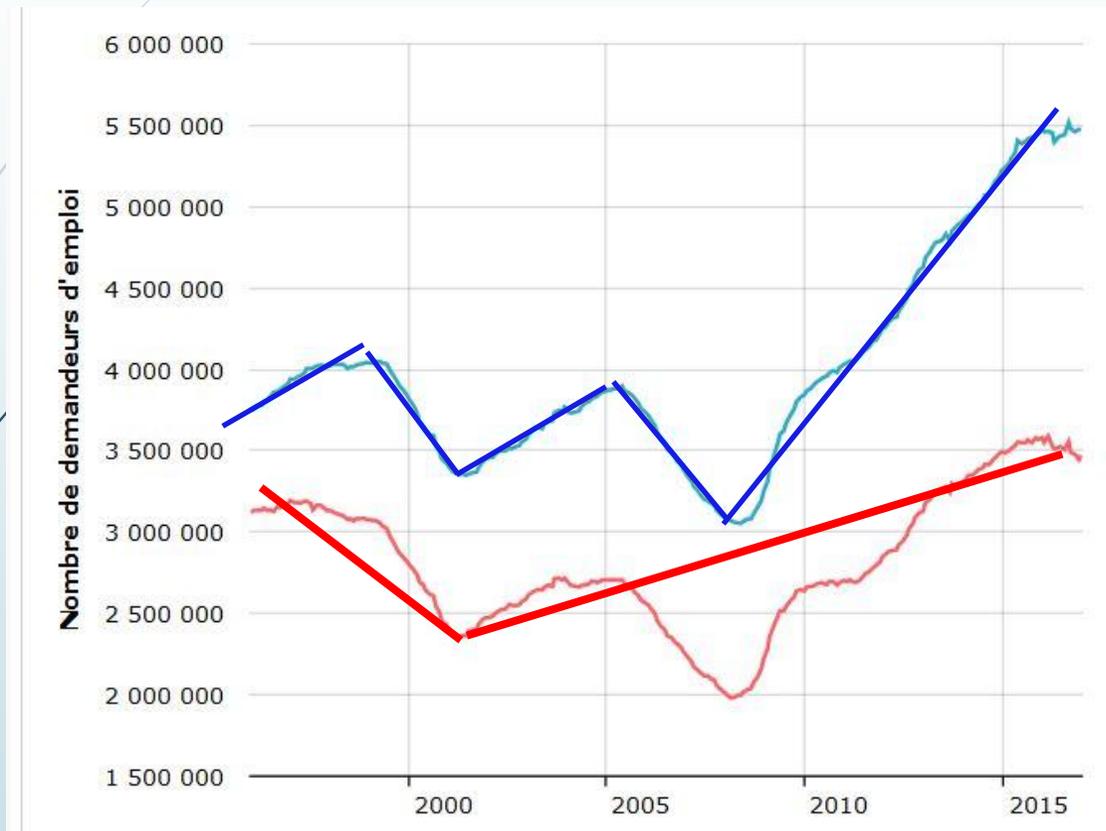
Disraeli/Twain?

Méthodes statistiques



Données

Méthodes statistiques



Approximation

Méthodes statistiques & textes

Tableau lexical entier (TLE)

Alice
occurrences 27 904
formes 2 951

	Ch1	Ch2	Ch3	Ch4	Ch5	Ch6	Ch7	Ch8	Ch9	Ch10	Ch11	Ch12
0 the	90	92	95	121	109	138	146	147	135	163	168	155
1 and	61	75	58	101	55	76	53	92	58	72	53	60
2 to	75	61	56	78	66	80	43	68	72	48	40	52
3 a	52	46	46	70	58	72	48	54	52	55	31	35
4 I	35	73	37	64	87	45	63	19	44	41	33	23
5 it	66	35	36	61	43	72	52	47	37	27	24	32
6 she	75	69	23	69	41	52	28	48	34	14	22	40
7 of	43	37	28	58	40	43	31	47	44	35	48	58
8 said	8	12	32	14	51	47	58	43	57	49	40	50
9 Alice	28	24	23	31	35	43	50	39	52	30	16	23
10 you	19	26	28	17	43	32	48	19	34	60	16	22
11 in	25	35	23	34	27	35	29	37	31	26	24	34
12 was	53	35	20	34	28	41	25	42	24	12	29	12
13 that	27	24	12	32	25	31	21	25	33	14	23	22
14 her	26	28	12	28	25	13	7	31	19	12	9	36
15 as	14	20	15	27	31	25	23	19	18	15	15	23
16 t	5	16	9	20	22	18	28	19	29	18	13	19
17 at	11	9	9	25	20	20	24	28	12	10	23	14
18 s	8	14	8	22	7	18	23	28	24	19	7	24

TLE :
2 951 lignes
13 colonnes
38 363 cases

Analyse statistique des tableaux lexicaux

approche locale
analyse des **spécificités**

approche globale
analyse factorielle

de
la
les
et
des

y

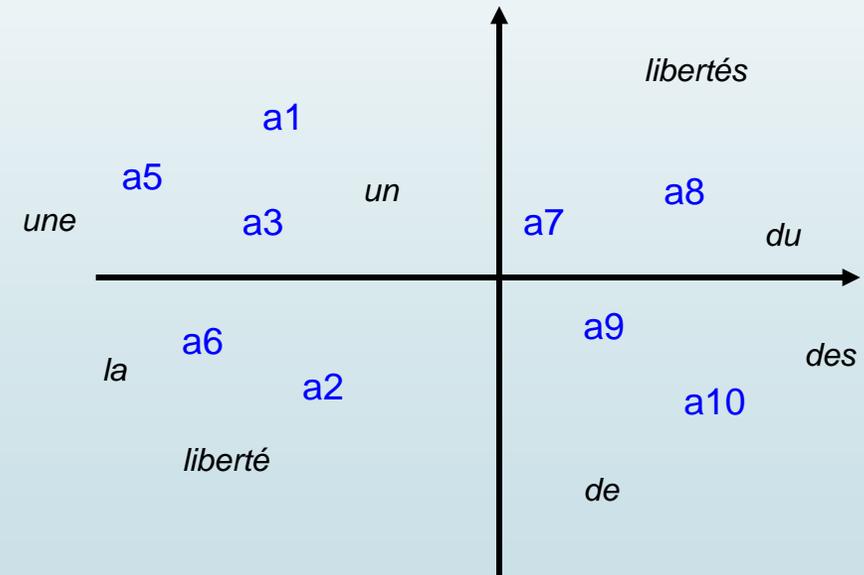
	a1	a _j	a3	a _n	
	89	99	82	76	f ₁
	78	58	58	63	f ₂
	35	45	35	38	f ₃
	28	28	24	28	f _i
	15	k _{ij}	20	20	
	1	0	0	0	
	t ₁	t _j ..		t _n	T

Analyse statistique des tableaux lexicaux

approche locale
analyse des **spécificités**

	a1	a _j	a3	a _n	
de	89	99	82	76	f ₁
la	78	58	58	63	f ₂
les	35	45	35	38	f ₃
et	28	28	24	28	f _i
des	15	k_{ij}	20	20	T
	1	0	0	0	
	t ₁	t_j		t _n	

approche globale
analyse factorielle



Analyse statistique des tableaux lexicaux

Méthode des spécificités 1

----- Chapter 1-----

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do : once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought **Alice** 'without pictures or conversation?'

...

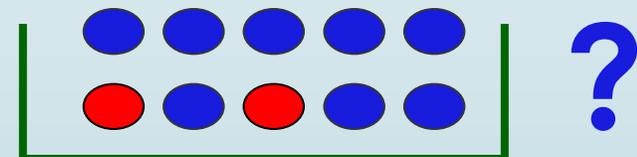
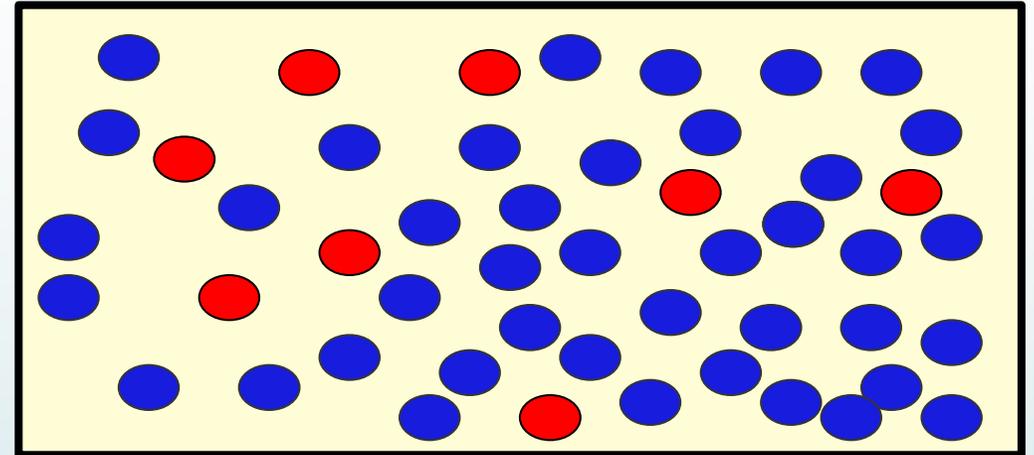
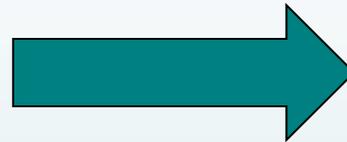
----- Chapter 2-----

'Curiouser and curiouser!' cried **Alice** (she was so much surprised , that for the moment she quite forgot how to speak good English) ; ' now I ' m opening out like the largest telescope that ever was !

/ ... / but I must be kind to them , ' thought

Alice

....

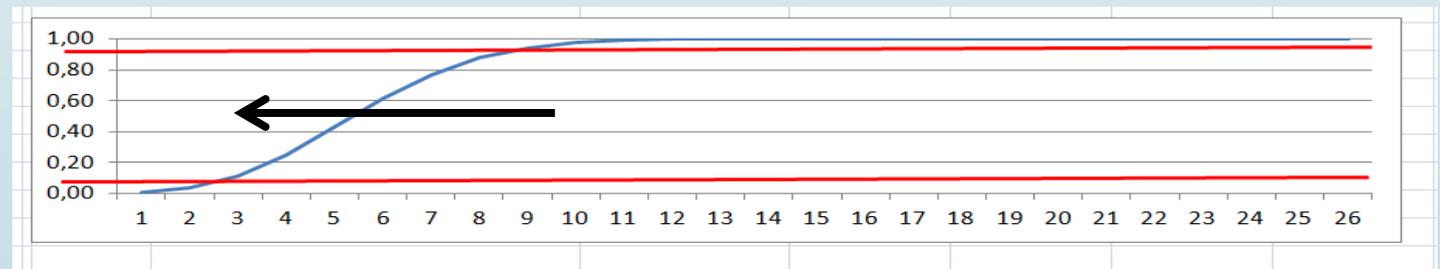
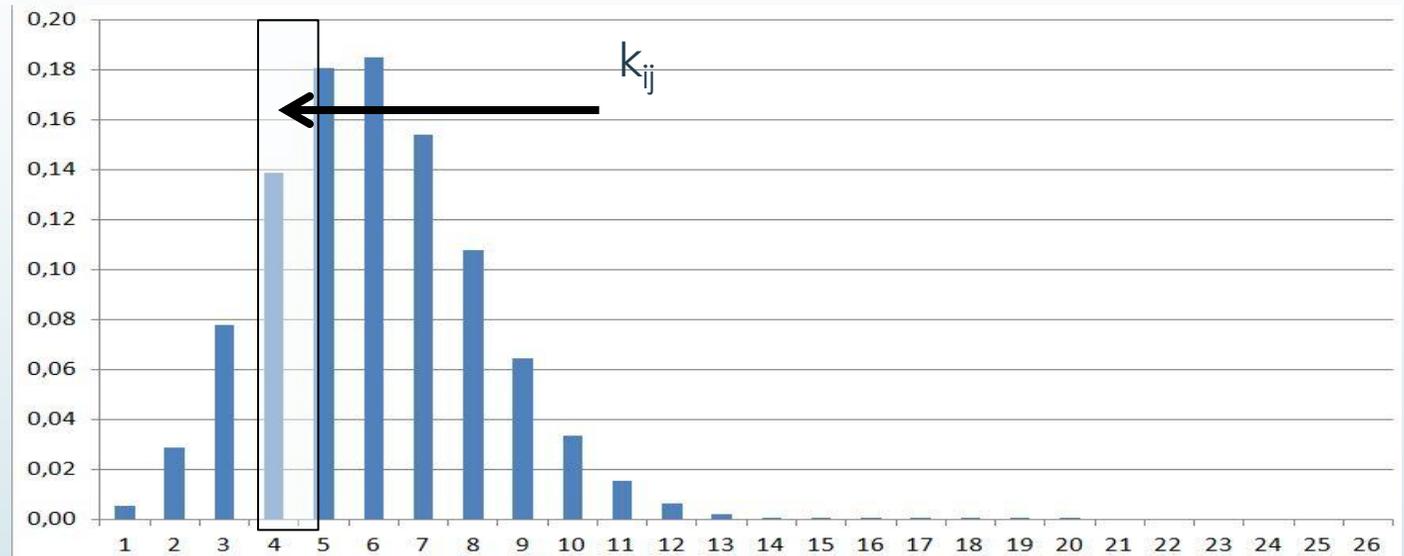


Analyse statistique des tableaux lexicaux

Méthode des spécificités 2

Paramètres de la loi
hypergéométrique :

T taille du corpus.
 t_j taille de la partie j
 F_i fréquence de la forme i .



Analyse statistique des tableaux lexicaux

Méthode des spécificités 3

formes spécifiques du chapitre 1

Segment	F	f	S
<u>she</u>	515	75	8
<u>was</u>	355	53	6
<u>it</u>	532	66	5
<u>down</u>	100	20	5
<u>eat</u>	17	6	4
<u>wonder</u>	19	6	4
<u>very</u>	138	23	4
<u>way</u>	57	13	4
<u>think</u>	53	11	4
<u>through</u>	13	5	4

The	113	2	-3
s	202	8	-3
the	1559	90	-4
t	216	5	-4
<u>said</u>	461	8	-9

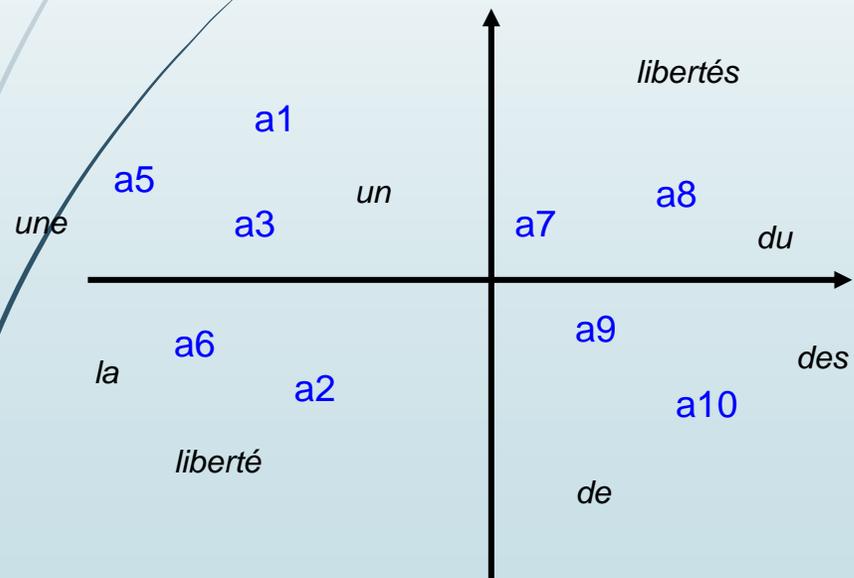
spécificités de la forme Alice dans les chapitres



Analyse statistique des tableaux lexicaux

Analyse factorielle des correspondances

approche globale



typologie à partir des proximités :

- a) entre les lignes du tableau
- b) entre les colonnes du tableau

	a1	a2	a3 ...	an	
de	89	99	82	99	76
la	78	58	58	58	63
les	35	45	35	45	38
et	28	28	24	28	28
des	15	k_{ij}	20	k_{ij}	20
y	1	0	0	0	0
	t_1	t_j	t_3 ...	t_n	f_1 f_2 f_3 f_4 f_i

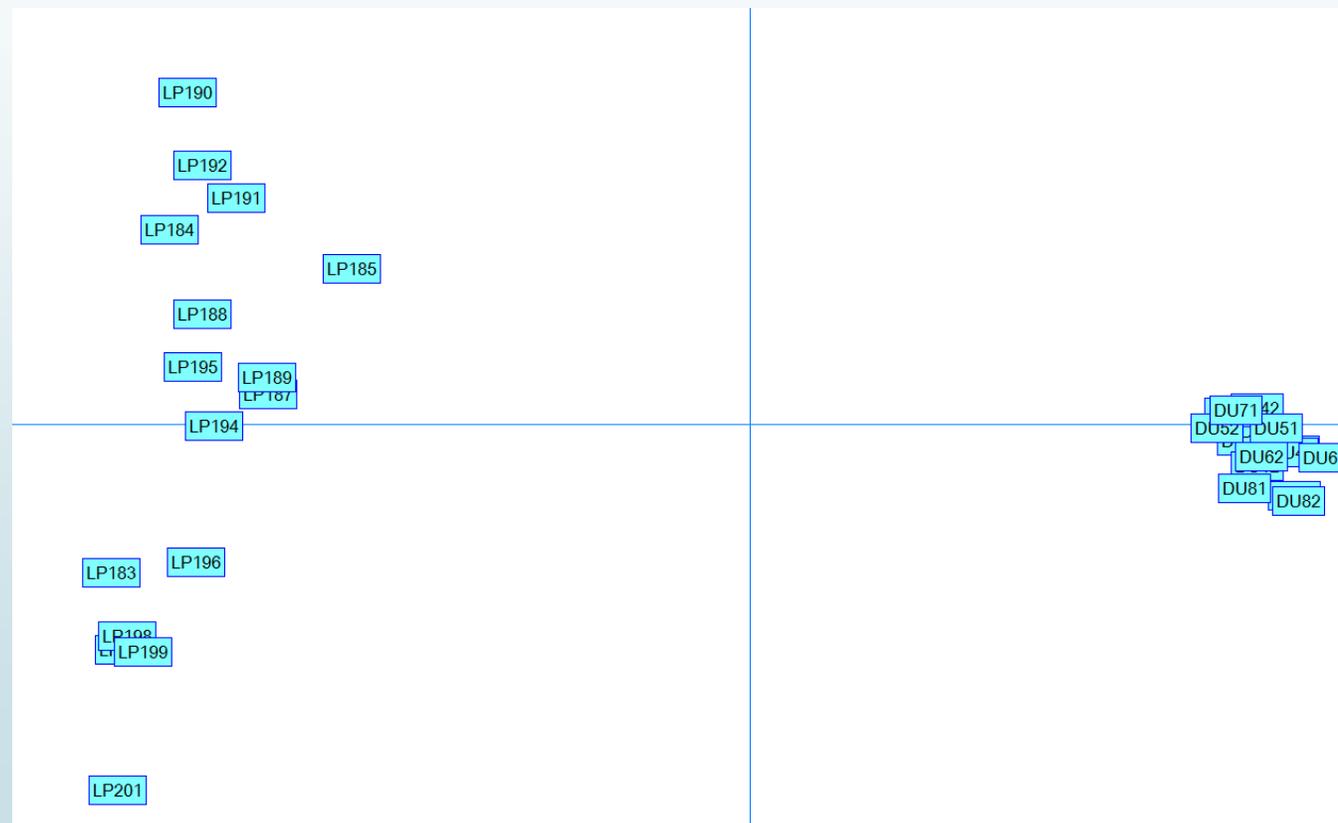
T

Analyse statistique des tableaux lexicaux

corpus composé de deux volets très différents

a) DU = *Le Père Duchesne* de Hébert (1793-1794)

b) LP = *Les français d'abord*, Le Pen (1984)



Top 5 textométrique des (plus mauvaises) idées reçues sur l'analyse textométrique



- Les analyses textométriques ne font que confirmer ce qu'une lecture attentive nous aurait permis de découvrir
- Il faut **impérativement** soumettre les textes à des traitements avant d'effectuer des comptages statistiques
- Les outils statistiques élémentaires (pourcentages, moyennes, etc.) suffisent dans toutes les situations
- Le sens des mots se trouve dans les dictionnaires
- Certains mots véhiculent du sens d'autres non

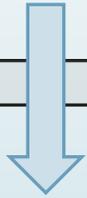


Séries textuelles chronologiques

Approches statistiques des corpus de textes

connaissances
a priori

→ 20e



données

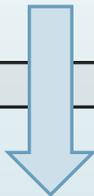
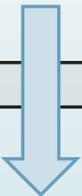
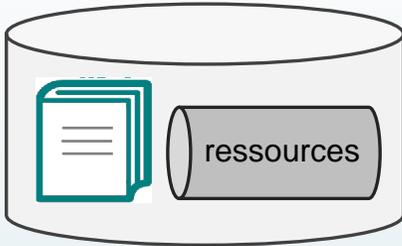
Approches statistiques des corpus de textes

connaissances
a priori

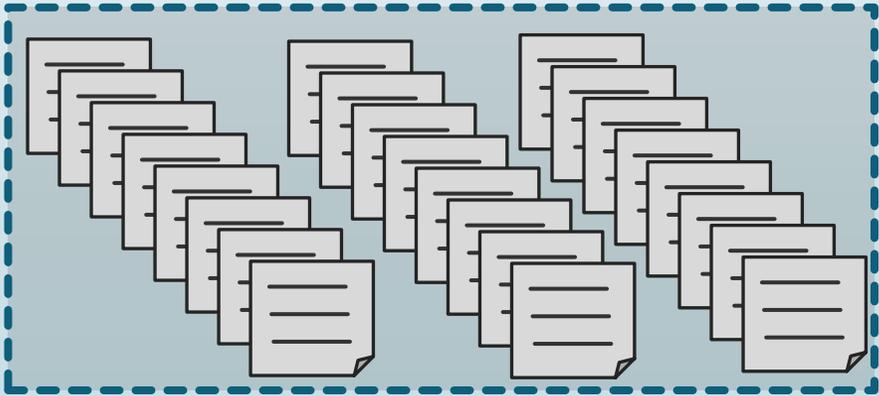
→ 20e



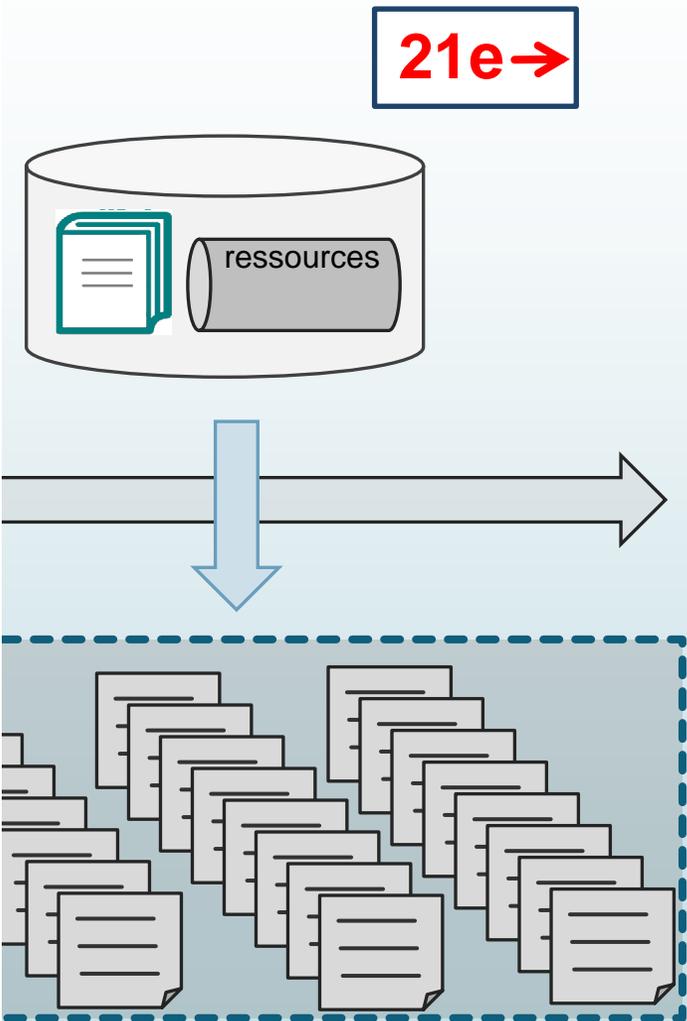
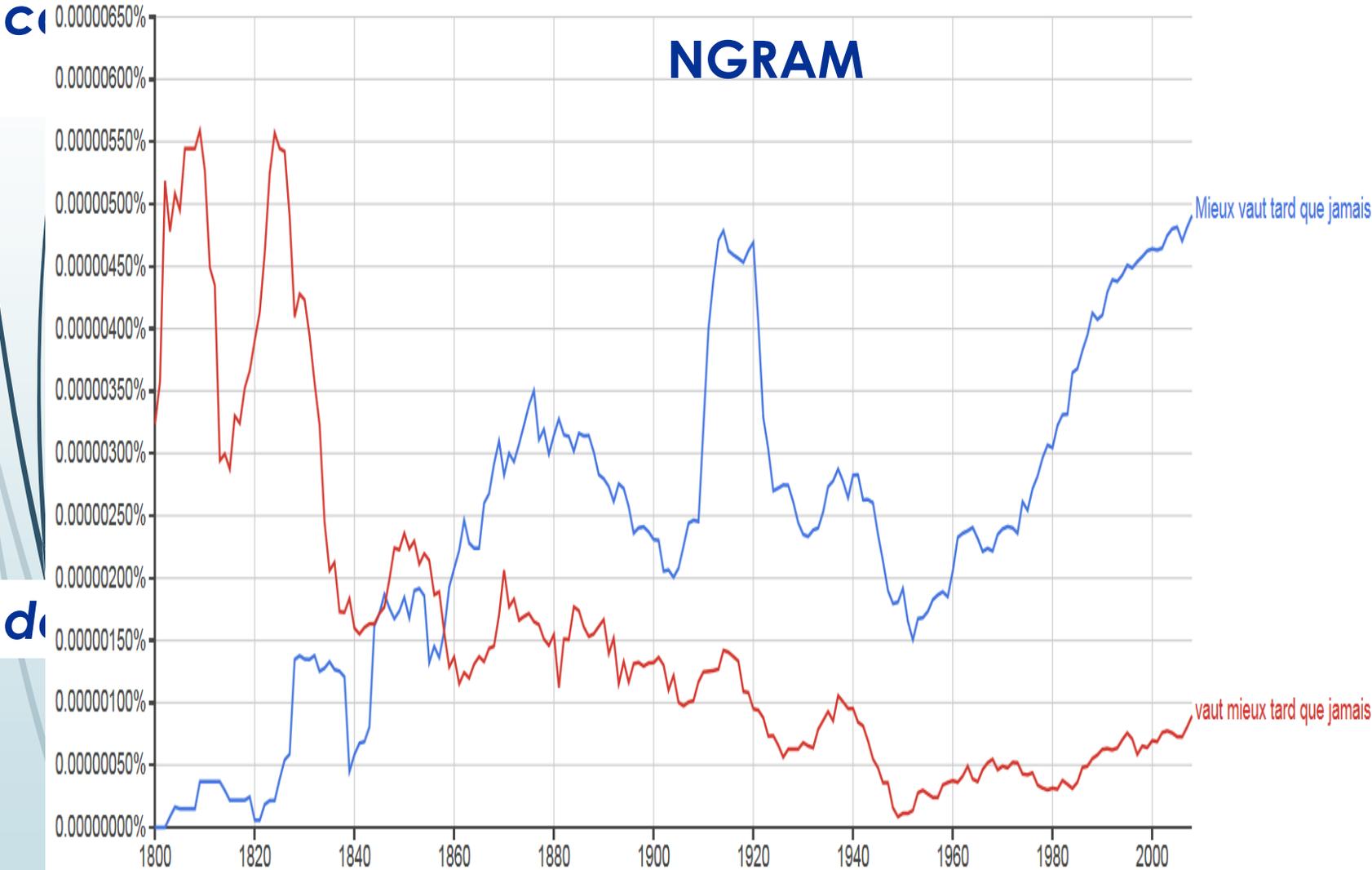
21e →



données



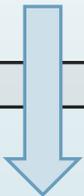
Approches statistiques des corpus de textes



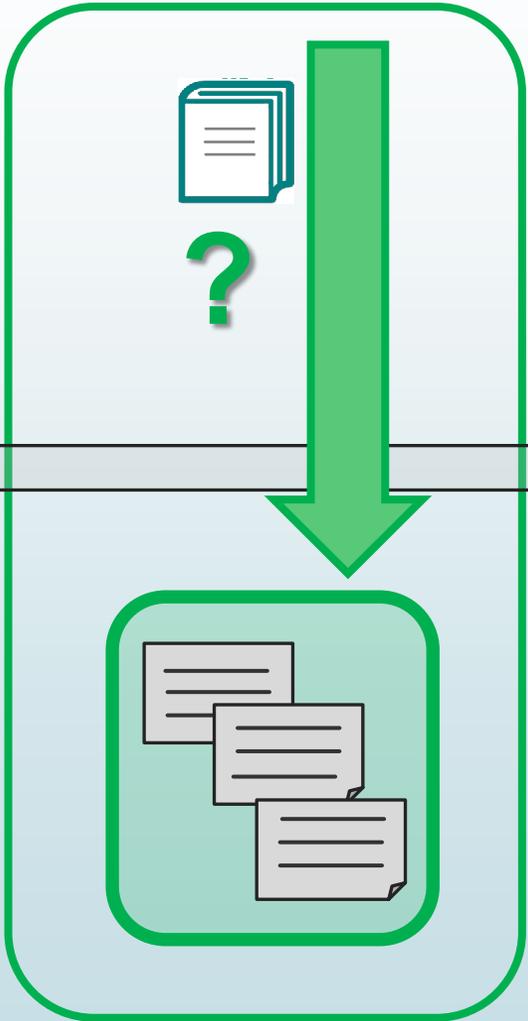
Approches statistiques des corpus de textes

connaissances
a priori

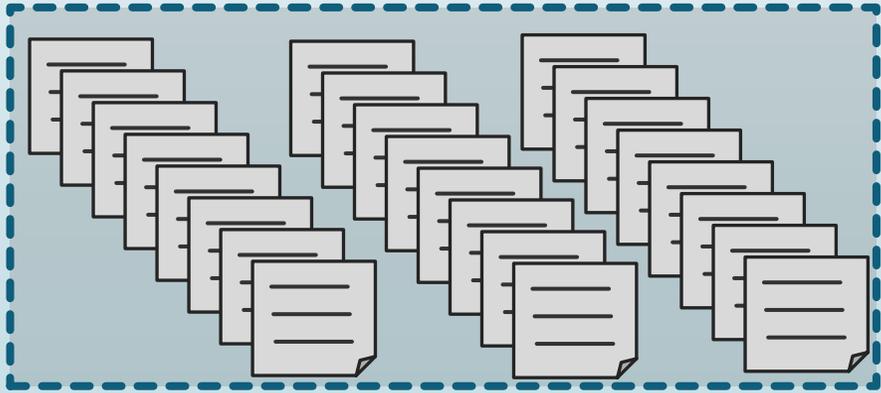
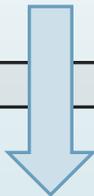
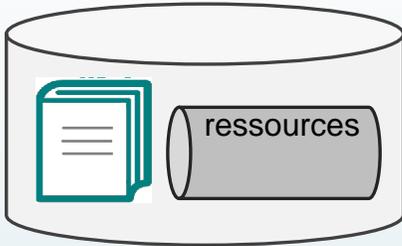
→ 20e



données



21e →



application areas : *historical studies*

1975 *Le Père Duchesne de Hébert* (1793-1794) (série textuelle chronologique)

tokens : 140 000
types : 11 000

<mois=01><numero=260> <Epg=1>

§ la grande douleur du *père *duchesne au sujet de la mort de *marat assassiné à coups de couteau par une garce du *calvados, dont l'évêque *fauchet était le directeur. ses bons avis aux braves *sans-culottes pour qu'ils se tiennent sur leurs gardes, attendu qu'il y a dans *paris plusieurs milliers de tondus de la *vendée qui ont la patte graissée pour égorger tous les bons citoyens.

§ *marat n'est plus, foutre. peuple, gémiss, <Epg=2> pleure ton meilleur ami; il meurt martyr de la liberté. c'est le *calvados qui a vomi le monstre sous les coups duquel il vient de périr. une jeune fille, ou plutôt une furie armée par les prêtres, et pénitente, dit on, du cafard *fauchet, part de *caen pour exécuter cet horrible attentat. elle arrive à *paris, et après avoir acheté un large couteau au



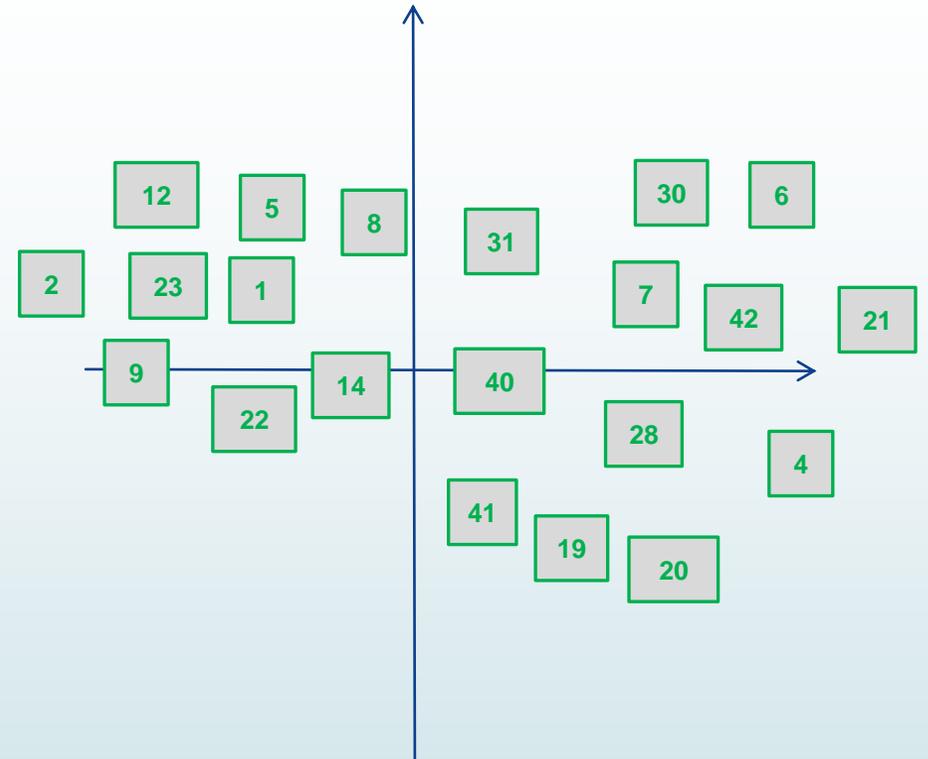
Je suis le véritable Père Duchesne, foutre !

LA GRANDE JOIE DU PÈRE DUCHESNE.

APRES avoir vu défiler la procession des Brissotins, des Girondins et des Rolandins, pour aller jouer à la main chaude à la place de la Révolution. Le testament de Cartouche Brissot, et la confession du prêtre Fauchet qui a fait le caffard jusqu'à la fin, pour faire pleurer les vieilles dévotes, mais qui, dans le fond du cœur, se foutoit autant du père éternel que du grand diable Belzébuc.

ADIEU paniers, vendanges sont faites; tous

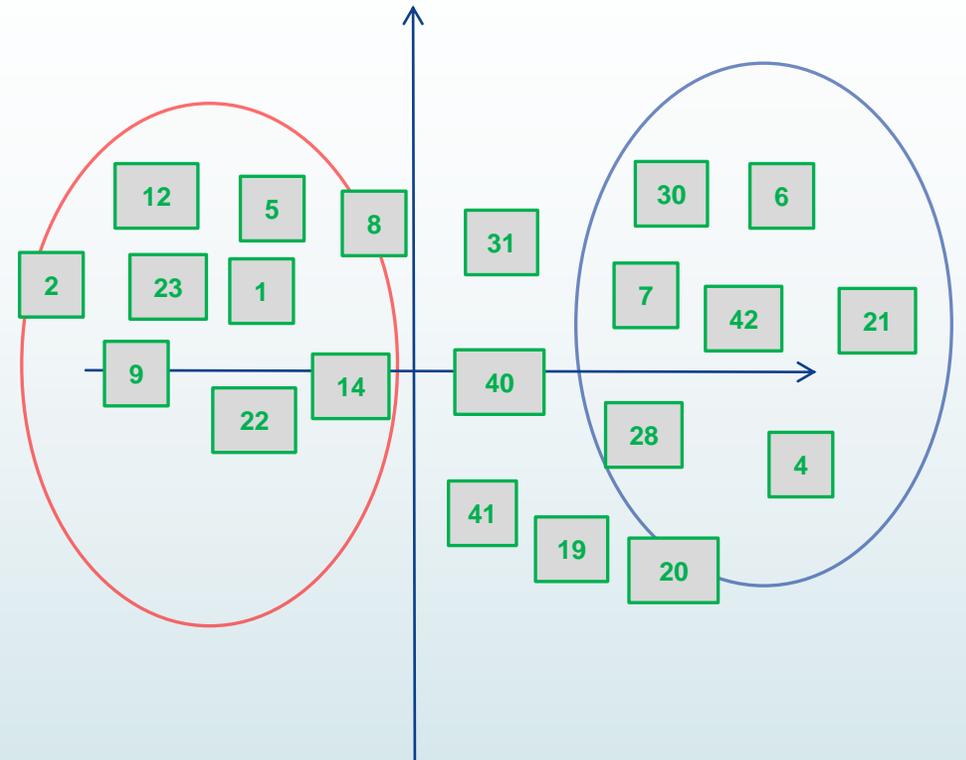
Déterminer une échelle d'analyse



numéro



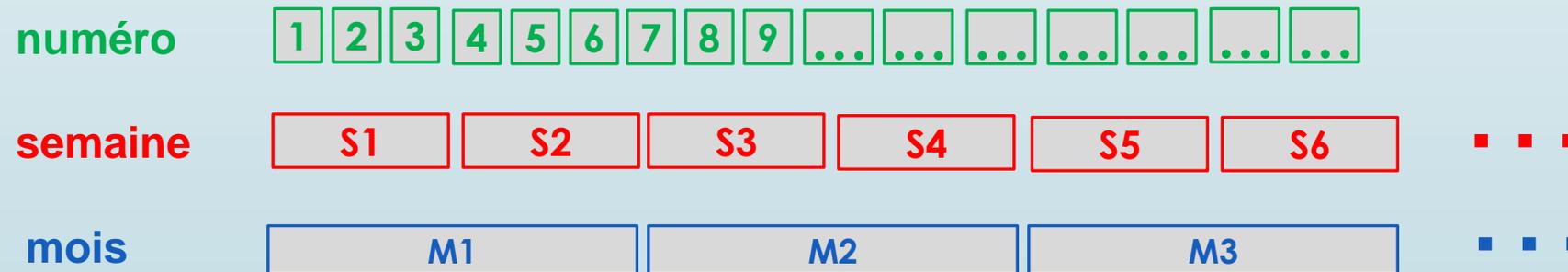
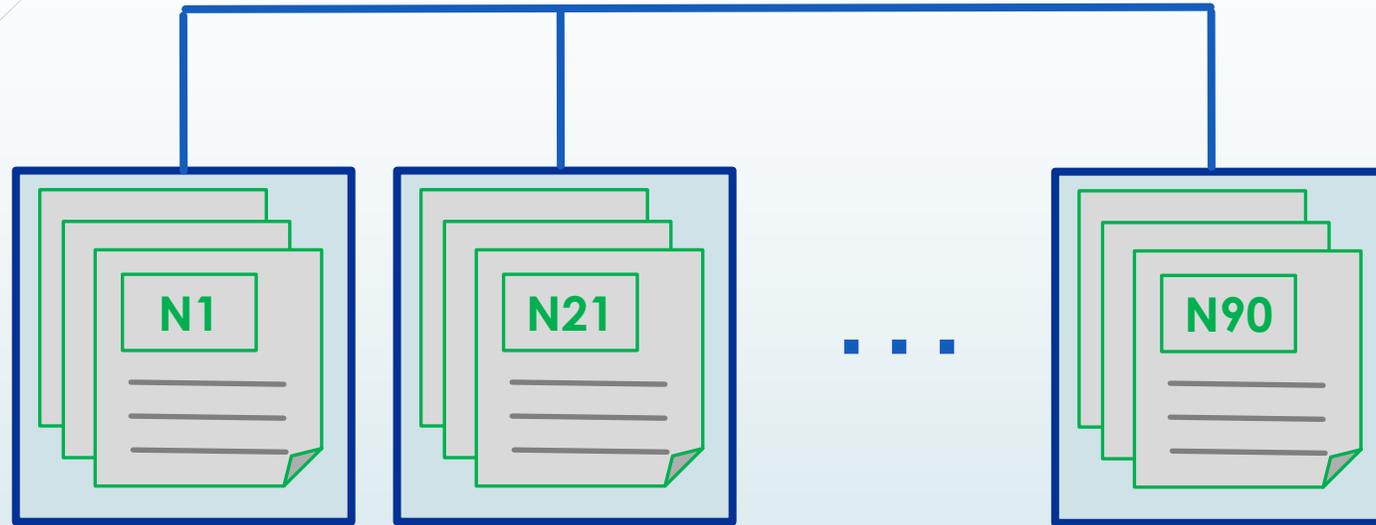
Déterminer une échelle d'analyse



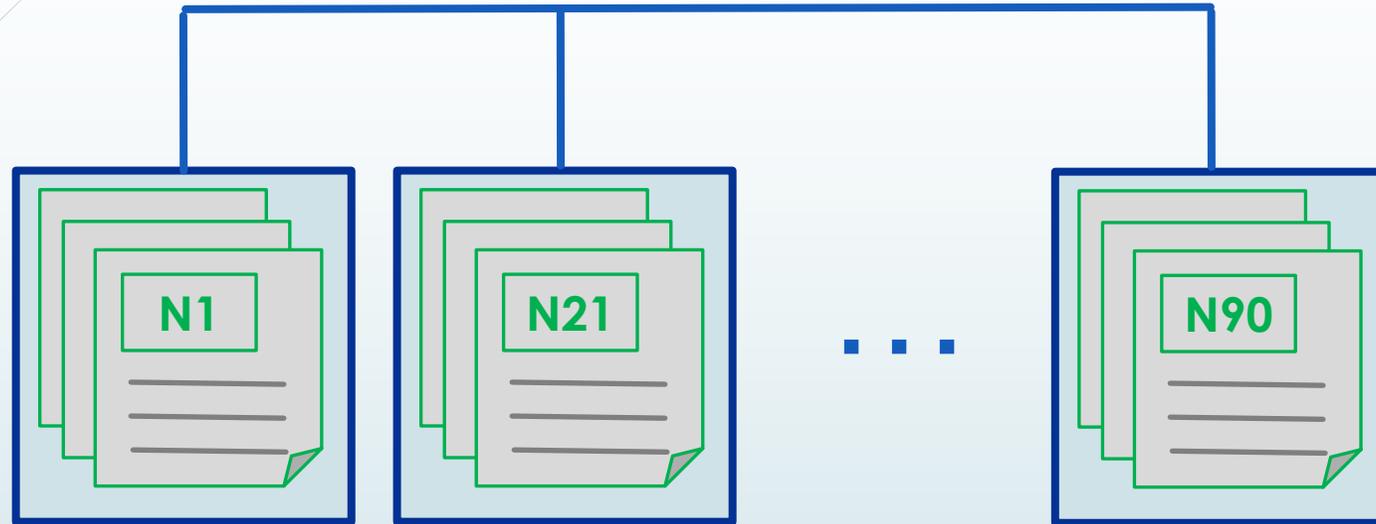
numéro



Déterminer une échelle d'analyse



Déterminer une échelle d'analyse



numéro 1 2 3 4 5 6 7 8 9

semaine S1 S2 S3 S4 S5 S6 ...

mois M1 M2 M3 ...



Déterminer une échelle d'analyse

Division en mois de parution



M1

M8

M2

M7

M3

M6

M4

M5

numéro



semaine

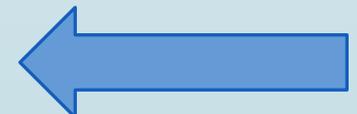


...

mois



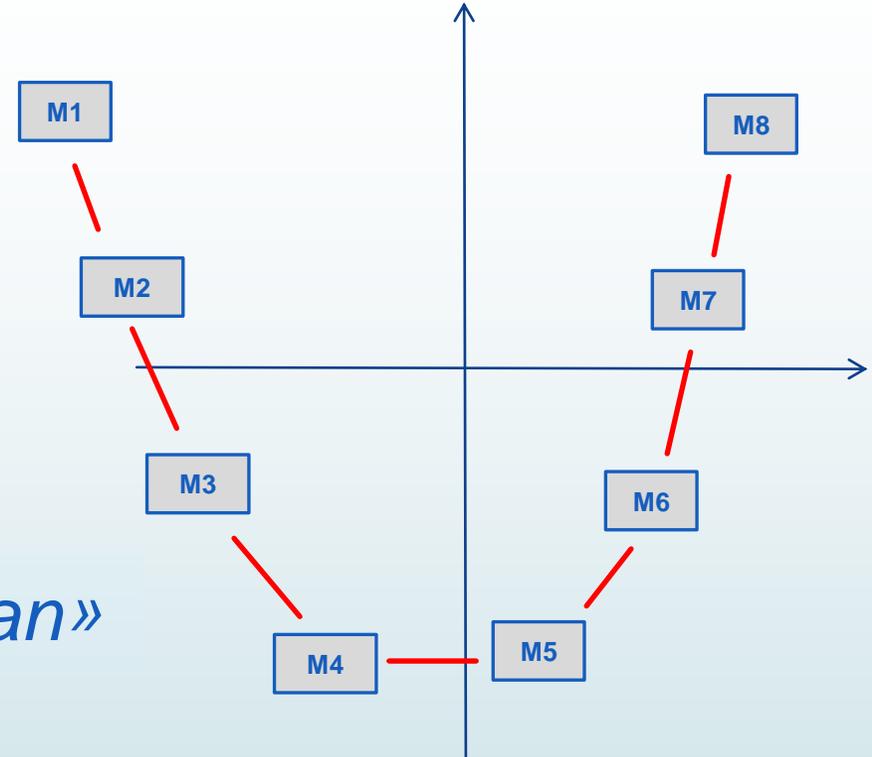
...



Déterminer une échelle d'analyse



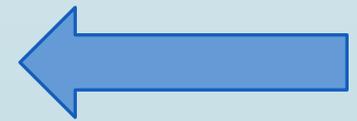
«effet Guttman»



numéro	1	2	3	4	5	6	7	8	9
semaine	S1	S2	S3	S4	S5	S6										
mois	M1	M2	M3													

...

...

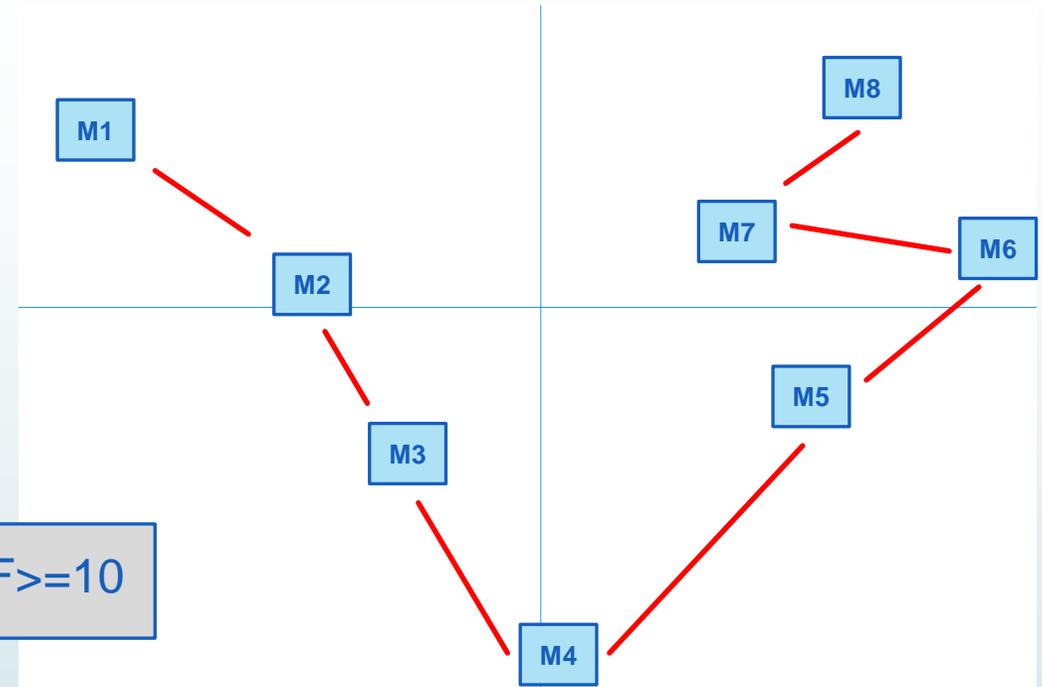


Utiliser la structure chronologique du corpus

Résultat sur le corpus Duchesne



Duchesne $F \geq 10$



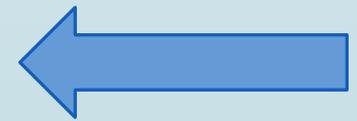
numéro 1 2 3 4 5 6 7 8 9

semaine S1 S2 S3 S4 S5 S6 ...

mois M1 M2 M3 ...

...

...

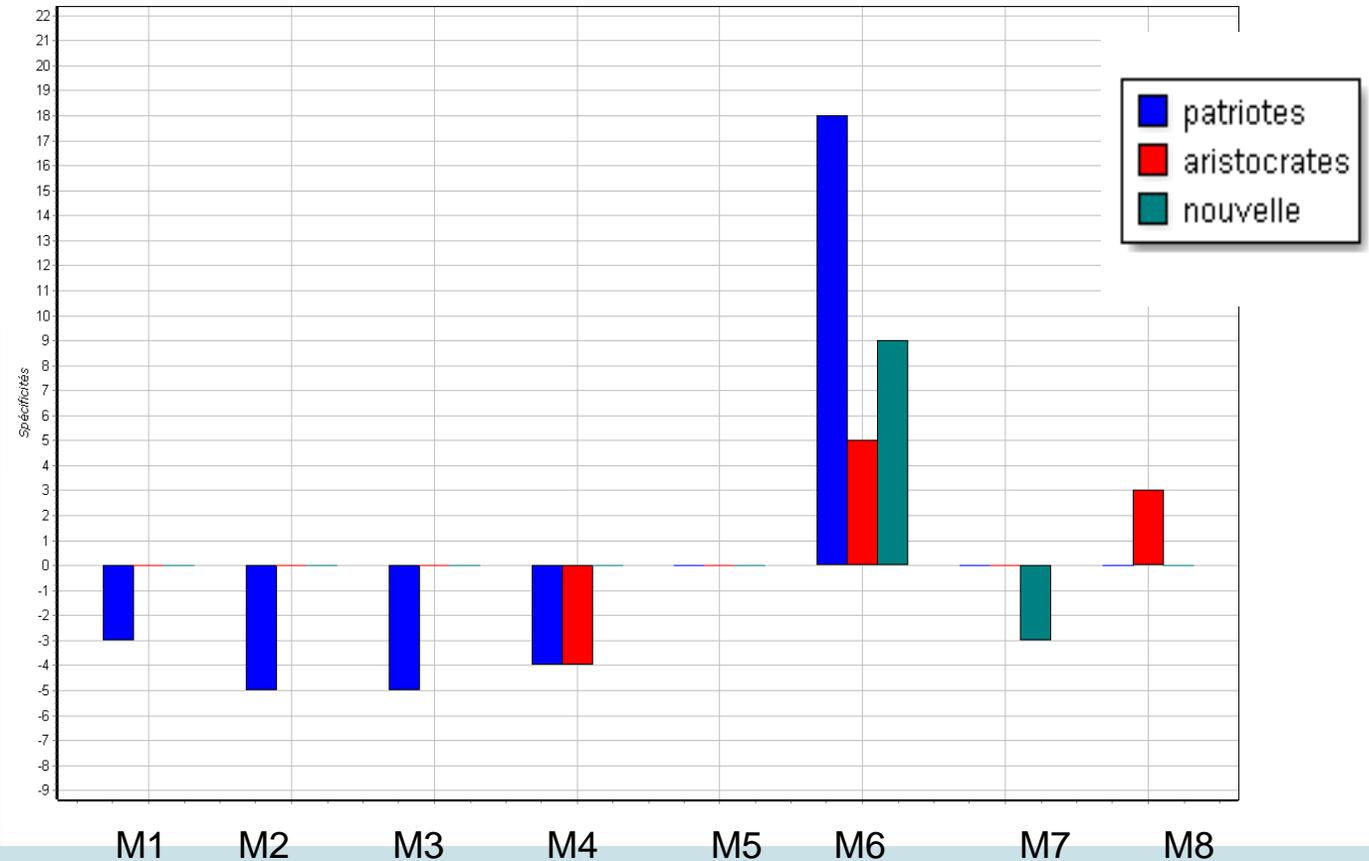


Utiliser la structure chronologique du corpus

spécificités de la partie 6

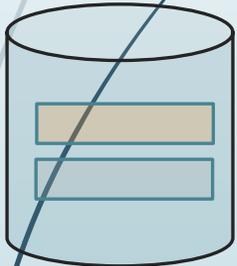
Terme	F ...	f	S ...
patriotes	152	60	18
*phélipotin	13	13	13
les patriotes	91	36	11
nouvelle	46	21	9
est là que	10	8	7
*vincent	15	10	7
les	4748	687	6
là que	12	8	6
est là	36	15	6
conspirateurs	48	18	6
ai	202	48	6
généraux *sans	10	7	5
contre	298	60	5
dis	34	13	5
aristocrates	133	32	5
diviser	16	9	5
mes	120	30	5
les meilleurs	31	13	5
accuser	10	7	5
comité	57	17	4
comité de salut ...	14	7	4

étudier des moments particuliers du corpus



Utiliser la structure chronologique du corpus

repérer des *significations locales*



-- Partie : M6, nombre de contextes : 21
*phélipoterie ; s ' il propose de former une
nt recruter tous les brigands et former une
re est venu à son secours . " formons une
à la chandelle ; je savais qu ' il existait une
le colère du *père *duchesne . de voir une
ongtemps que je furetais pour découvrir la
les courriers extraordinaires ont apporté la
ste plus . la convention vient de recevoir la
s bonnets rouges . ces *sans - culottes de
' il débite gratis . au dénouement de cette
es sont unis , ne sera pas le jouet de cette
s ce n ' est pas assez , il faut savoir quelle
la *vendée , ou plutôt pour en former une
es trahisons , serait le connétable de cette
eurs desseins . quand je découvre quelque
crier : victoire , vive la république . grande
ommères , en réjouissance de cette bonne
sur parole . voilà , mon vieux , une grande
couronnés , quand ils vont apprendre cette
aordinaires ont apporté la nouvelle d ' une
le la détruire , lui donnent au contraire une

nouvelle *vendée en mettant en liberté ses bons
nouvelle *vendée en proposant d ' ouvrir les pris
nouvelle *vendée , lui a dit l ' animal cornu , form
nouvelle clique d ' aristocrates , de modérés , d '
nouvelle clique de modérés , de feuillants , d ' ari
nouvelle conspiration qui est maintenant sur le ta
nouvelle d ' une nouvelle victoire , on a annoncé
nouvelle de la destruction totale des brigands ; ce
nouvelle fabrique se sont répandus dans les sect
nouvelle farce contre - révolutionnaire , nous déc
nouvelle intrigue . après avoir sauvé la *france ,
nouvelle liste civile paye les rhapsodies qu ' il dék
nouvelle , mais suivons pied à boule notre endon
nouvelle majesté égorgeante ; *brissot serait son
nouvelle manigance contre les *sans - culottes ,
nouvelle , mes amis , leur ai je dit ; femmes , tais
nouvelle , qui donne le coup de grâce aux aristoc
nouvelle qui va faire gueuler , dans tous les coins
nouvelle . sa grande colère contre certains jean -
nouvelle victoire , on a annoncé que les lignes de
nouvelle vie . ils tiennent les patriotes en haleine

Utiliser la structure chronologique du corpus

dégager le vocabulaire de l'évolution

Terme	F	f	S	Parties	Tresse chronologique
nobles	42	21	+9	M1	—————
*marat	65	27	+9	M1	—————
elle	368	278	+9	M1	—————
la constitution	40	31	+11	M1 - M2	————— —————
*custine	70	47	+12	M1 - M2	————— —————
constitution	72	53	+16	M1 - M2	————— —————
*paris	171	93	+14	M1 - M2	————— —————
vous	1097	705	+13	M1 - M4	————— ————— ————— —————
est vous qui avez	10	10	+10	M4	—————
tu	296	200	+9	M3 - M6	————— ————— ————— —————
on	856	288	+9	M5 - M6	————— —————
*phélipotin	13	13	+13	M6	—————
nouvelle	46	21	+9	M6	—————
les patriotes	91	36	+11	M6	—————
patriotes	152	60	+18	M6	—————
*vincent	15	14	+9	M6 - M7	————— —————
républicains	129	91	+9	M5 - M8	————— ————— ————— —————
foutre	1100	784	+10	M7 - M8	————— —————
*courtille	13	10	+9	M8	—————

Site d'annonces pour le Séminaire (ED268) :

Analyse du discours et textométrie / Statistiques textuelles
Resp. André Salem

(mise à jour : **14/12/2017**)

Attention : changement d'horaire ** 15h30-17h30 **

Dates des prochains cours :

les cours auront lieu les : mercredi 24 janvier, 7 et 21 février, 7 et 28 mars, 11 avril
de 15h30-17h30, salle Brunot, Ilpga RdC, 19 rue des Bernardins, 75005

un stage dont la date sera fixée avec les participants est prévu en juin 2018

des informations sur le séminaire sont disponibles sur le site :

<http://www.tal.univ-paris3.fr/as/sp3/>

Documents pour le cours

- [Statistiques textuelles \(ouvrage\)](#)
- [Diapositives du cours](#)
- Logiciel [Lexico5.5.1](#)
- Logiciel [Lexico3.6](#)
- [Tutoriel1](#)
- [Tutoriel2](#)
- [Explorations textométriques](#)

contact : salem@msh-paris.fr Andre.Salem@univ-paris3.fr



Merci pour votre attention

Je suis prêt à répondre à vos questions, si vous en avez