

*Table-ronde autour de l'ouvrage*  
**La sémantique au pluriel**  
**Théories et méthodes**  
*Séminaire Céditec - ConSciLa*  
*Université Paris-Est Créteil, 10 mars 2023.*



# Sémantique textométrique

Bénédicte Pincemin

Université de Lyon, CNRS, IHRIM UMR5317



This work is licensed under the Creative Commons Attribution 4.0 International License.  
<http://creativecommons.org/licenses/by/4.0/>

# Plan

1. Situer la textométrie
2. Illustrer concrètement ses méthodes
3. Pointer ses liens à la réflexion sémantique

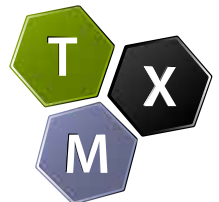
# La sémantique au pluriel.

## Théories et méthodes

- La textométrie n'est pas une théorie sémantique.
- Elle apporte un éclairage expérimental sur le fonctionnement de la langue.
- Elle ne correspond pas à une théorie particulière.

# La textométrie : concrètement ?

- Des outils logiciels pour observer et relire un corpus, de façon globale, systématique, réfléchie.
  - Traitements **quantitatifs** (fréquences, statistiques) et
  - Fonctionnalités **qualitatives** (lecture, annotation)
- Positionnement original dans les Humanités numériques
  - Pas assimilable au *Text mining* ou à l'IA.



# Panorama de méthodes textométriques

## TEXTES

*Contextes, cotextes  
(locuteur, période...)*

« contenant »



## MOTS

*Unités, motifs, traits...*

« contenu »

Les méthodes articulent  
des textes et des mots (au sens large),  
des « contenant » et des « contenu »,  
elles **s'inscrivent dans un continuum**  
entre ces deux pôles.

# Panorama de méthodes textométriques

## TEXTES

*Contextes, cotextes  
(locuteur, période...)*  
« tenants »



## MOTS

*Unités, motifs, traits...*  
« contenus »

**Spécificités**

**Retour au texte**

Lecture, édition du  
texte, visualisation du  
document source

**Classification**

appliquée aux textes,  
analyse arborée

**AFC**

analyse factorielle des  
correspondances

Évolution  
spécificités  
chronologiques

**Concordance**

**Méthode ALCESTE**

méthode Reinert

Déroulement  
progression,  
topologie

**Cooccurrences**

**Segments répétés**

inventaire distributionnel

**Vocabulaire**

Analyse des  
similitudes  
réseaux de mots

Etc.

# Corpus illustratif tiré du livre

- Comme corpus partagé par l'ensemble des participants au séminaire, nous avons pensé à une première exploration du livre *La sémantique au pluriel* et des 20 chapitres qui le composent.
- Cela a été possible grâce à une version numérique de l'ouvrage que Dominique Ducard a pu nous communiquer spécifiquement pour les besoins de cette communication. Qu'il en soit très chaleureusement remercié !
- Détails techniques pour ceux qui voudraient passer le nez dans l'arrière-cuisine :
  - La source PDF a été convertie en texte brut avec conservation des sauts de page (macro TXM prototype PDF2TXT) puis éditée à l'aide de l'outil oXygen pour en faire un fichier type TEI, avec identification des différents chapitres. Un soin particulier a été apporté au traitement des mots coupés par une césure en fin de ligne, pour qu'ils puissent être bien reconnus. Ce fichier a ensuite été importé dans TXM, puis aussi IRaMuTeQ. Le corpus obtenu compte près de 220 000 mots.
  - Une autre voie possible aurait été d'utiliser le fichier DOCX tiré du PDF, de le subdiviser en autant de fichiers que de chapitres, et d'utiliser un import dédié au traitement de texte.



# Panorama de méthodes textométriques

## TEXTES

*Contextes, cotextes  
(locuteur, période...)  
« contenant »*



## MOTS

*Unités, motifs, traits...  
« contenus »*

**Spécificités**

**Retour au texte**

Lecture, édition du  
texte, visualisation du  
document source

**Classification**

appliquée aux textes,  
analyse arborée

**AFC**

analyse factorielle des  
correspondances

**Cooccurrences**

**Vocabulaire**

**Segments répétés**

inventaire distributionnel

**Concordance**

**Méthode ALCESTE**

méthode Reinert

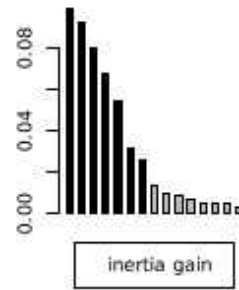
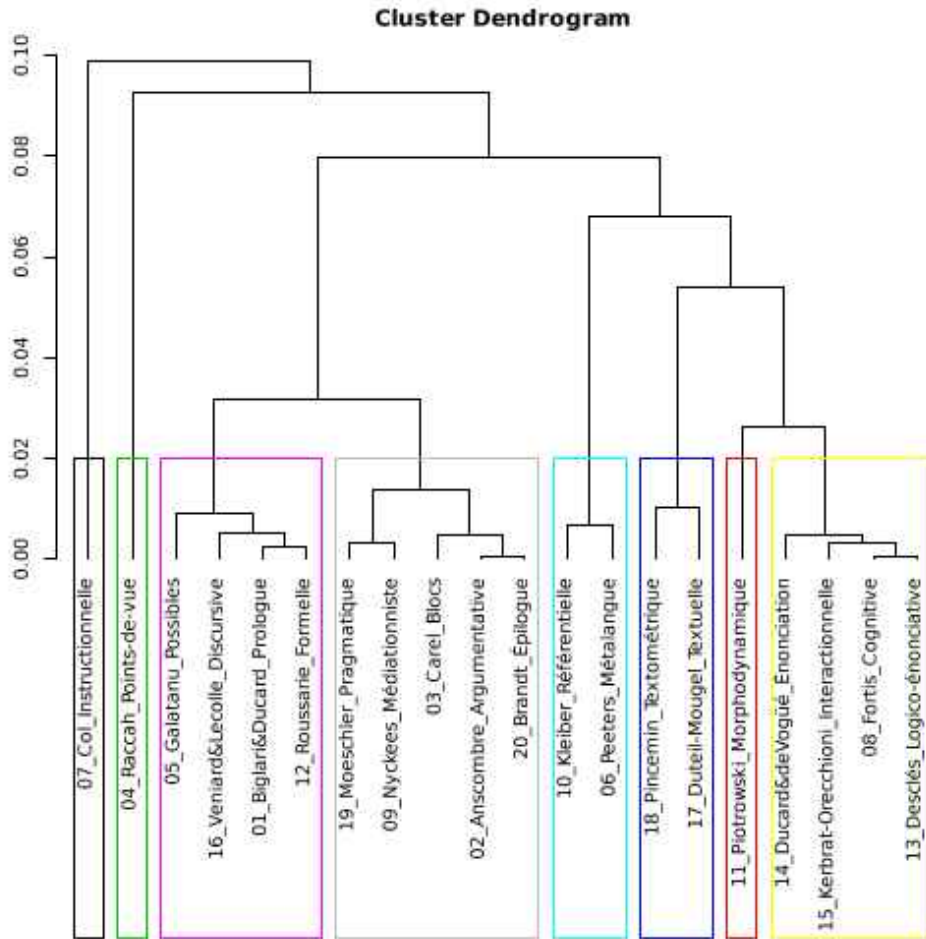
Regroupement des **textes en fonction de leur similarité**,  
visualisé souvent sous forme d'arborescence



# Illustration sur le corpus des 20 chapitres de *La sémantique au pluriel*



Hierarchical clustering



Liste des 47 mots (fréq.  $\geq 150$ ) sous-jacents à l'analyse :

analyse	monde
approche	mot
autre	notion
cas	objet
chose	phrase
construction	point
contexte	principe
différent	question
discours	rapport
discursif	relation
effet	représentation
élément	sémantique
énoncé	sens
exemple	signe
expression	signification
forme	sujet
interprétation	terme
langage	texte
langue	théorie
lexical	type
linguistique	unité
locuteur	valeur
même	vue
modèle	

**Classification ascendante hiérarchique des chapitres**  
*tels que décrits par les noms et adjectifs (lemmatisés)*  
*les plus fréquents (Fmin = 150, 20 textes, 47 mots)*

# Panorama de méthodes textométriques

## TEXTES

*Contextes, cotextes  
(locuteur, période...)*  
« tenants »



## MOTS

*Unités, motifs, traits...*  
« contenus »

**Spécificités**

**Retour au texte**

Lecture, édition du  
texte, visualisation du  
document source

**Classification**

appliquée aux textes,  
analyse arborée

**AFC**

analyse factorielle des  
correspondances

**Cooccurrences**

**Vocabulaire**

**Segments répétés**

inventaire distributionnel

**Concordance**

**Méthode ALCESTE**

méthode Reinert

Visualisation spatiale (comme une carte géographique)  
des **proximités des textes et des mots** :  
configuration d'ensemble, **contrastes dominants** au sein du corpus



# Panorama de méthodes textométriques

## TEXTES

*Contextes, cotextes  
(locuteur, période...)*  
« contenant »



## MOTS

*Unités, motifs, traits...*  
« contenus »

**Spécificités**

**Retour au texte**

Lecture, édition du  
texte, visualisation du  
document source

**Classification**

appliquée aux textes,  
analyse arborée

**AFC**

analyse factorielle des  
correspondances

**Cooccurrences**

**Vocabulaire**

**Segments répétés**

inventaire distributionnel

**Concordance**

**Méthode ALCESTE**

méthode Reinert

À tout moment dans l'analyse,  
voir le **mot dans son contexte original**,  
feuilleter les textes, visualiser les documents sources

# Illustration sur le corpus des 20 chapitres de *La sémantique au pluriel*

SEMANTIQUES22-230308/<[frlemma="discour... ✕

Query [frlemma="discours"]

ref	Left context	Pivot	Right context
p. 073, Carel, Blocs	nvaincant et tel autre non, tel	discours	banal et tel autre s
p. 073, Carel, Blocs	s -mêmes, la solidité de notre	discours	, sa qualité stylisti
p. 073, Carel, Blocs	istique. La langue arbitre nos	discours	.
p. 075, Raccah, Points	de vue que les énoncés et les	discours	amènent à adopte
p. 076, Raccah, Points	on du sens des énoncés et des	discours	concernent princip
p. 076, Raccah, Points	voir ce dont ces énoncés et ces	discours	parlent. Cette expl
p. 076, Raccah, Points	la compréhension même des	discours	, et qui fonctionne
p. 076, Raccah, Points	supposés idéologiques de ces	discours	. Nous verrons ens
p. 078, Raccah, Points	Dès que l'on reconnaît que les	discours	visent à faire adme
p. 079, Raccah, Points	ent les langues et non plus les	discours	-, fondées sur des
p. 079, Raccah, Points	urs -, fondées sur des faits de	discours	observables, desc
p. 079, Raccah, Points	prédire de nouveaux faits de	discours	. Néanmoins, les d
p. 079, Raccah, Points	de discours. Néanmoins, les	discours	qu'une telle étude
p. 079, Raccah, Points	ix langues dans lesquelles ces	discours	sont proférés. Ma
p. 079, Raccah, Points	conception polyphonique du	discours	et du texte, était a
p. 079, Raccah, Points	ernaient pas la langue mais le	discours	et les textes 7. À ce
p. 080, Raccah, Points	on, s'il est incontestable qu'un	discours	ne peut pas revenc
p. 080, Raccah, Points	ion erronée entre l'objet et le	discours	sur l'objet. En effet
p. 080, Raccah, Points	tives n'implique pas que tout	discours	sur ces entités soit
p. 080, Raccah, Points	subjectives peut constituer un	discours	objectif. Sémantiq
p. 080, Raccah, Points	es sur les interprétations des	discours	en situations. Sur c
p. 081, Raccah, Points	n situation. L'observation des	discours	en situation donne

1 - 100 / 257

semantiques22 - 77 ✕

- 76 -

Cette distinction, qui peut paraître une évidence, n'est pourtant faite que par un ensemble de courants, regroupés sous la dénomination de « sémantiques instructionnelles » les autres courants n'accordent, au mieux, qu'une différence de degré de généralité entre sens et signification. On verra aussi que cette distinction catégorique, outre son intérêt pour maintenir une sorte d'hygiène conceptuelle, est indispensable pour des raisons empiriques : ce qu'un sujet parlant comprend n'est accessible qu'à lui-même et n'est donc pas un observable scientifique ; ce qui, dans la langue, le lui fait comprendre est, en revanche, nécessairement accessible à tous : si ce n'était pas le cas, aucune langue humaine ne pourrait être acquise.

Cette distinction étant faite, entre le sens (non scientifiquement observable) que les individus construisent et les instructions et les contraintes (la signification) que les langues humaines imposent à sa construction, nous explorerons l'idée selon laquelle ces instructions et contraintes que les langues imposent à la construction du sens des énoncés et des discours concernent principalement la manière de voir ce dont ces énoncés et ces discours parlent. Cette exploration, ce point de vue sur une sémantique de l'argumentation (voir, notamment, Ducrot, 1988) qui développe des arguments pour une sémantique des points de vue (SPV : voir notamment Raccah, 2002 ; Carel et Ducrot, 2009 2 ; Carel, 2011a), amène à considérer l'orientation argumentative des énoncés comme une des manifestations de cette manière de voir, partiellement déterminée par la langue dans laquelle ces énoncés sont proférés : on sait déjà, depuis Aristote, qu'un argument n'est argument qu'en vertu de certaines conceptions préexistant à l'énoncé de l'argument. Nous verrons dans cette section que ceux des points de vue qui sont nécessaires à la compréhension même des discours, et qui fonctionnent donc comme des présupposés linguistiques, constituent aussi les présupposés idéologiques de ces discours. Nous verrons ensuite qu'une telle approche, dans laquelle les unités des langues fournissent des contraintes sur les points de vue que leurs usagers doivent adopter (même provisoirement) pour construire un sens pour les énoncés qui utilisent ces unités, permet de rendre compte d'un grand nombre de phénomènes liés à la construction du sens, et, en particulier, des phénomènes considérés comme relevant de l'argumentation dans la langue, dont cette approche peut être considérée comme une extension. On verra aussi qu'une telle approche, associant les implicites des argumentations à des présupposés idéologiques, permet de rendre compte de cette faculté extrêmement robuste que les sujets parlants ont de comprendre très vite, et avec très peu de matériau, à quelles idéologies les locuteurs se présentent comme adhérant.

Page 77 / 475

Concordance du mot discours, et **retour au texte** (page 76).

*La présentation du texte est relativement fidèle (restitution de la pagination et des lignes) mais dégradée (perte de la mise en forme des tableaux et graphiques, des alinéas, des mises en valeur typographiques, etc.). C'est en partie une question de temps de préparation, mais aussi de droits (renvoyer à l'ouvrage publié pour l'édition de qualité faisant référence.)* 13

# Illustrations de retour au document source dans d'autres contextes



OBLCUNEIF Lexique / OBLCUNEIF OBLCUNEIF:[word="ma"]

[-2-]

[ ]  
 adi te+em-ka la aš-pur-am ma  
 si-pi-ir l-im sa ih-he-ruú  
 la l-mu-ru-nim  
 mu-ú a-na si-íp-ri-im  
 ga-am-ri-im  
 la úš-ta-ar-du-ú  
 ú is-tu si-pi-ir l- sa i-na-an-na  
 sa-ab-ta-ti  
 i-na he-re-e-em ta-ag-dam-ru  
 (r)-UD-KIB-NUN-(ki) sa is-tu  
 UD-UNU-(ki)  
 a-di SES-UNU-(ki)  
 mi-iq-ti-sa ú-su-úh  
 ha-mi-ša su-ut-bi  
 su-te-še-er-si

[-2-]

[ ]

5

10

page OBLCUNEIF - translit ...

[ translit | cuneiform | facs ] 3 / 3 PDF Notice

Requête : [word="ma"] Chercher Réglages

Référence	Contexte gauche	Pivot	Contexte droit
216	TXM_cuneif_Ha_S_Abb_2_4	ta aš pur am um ma at ta	ma ERIN <sub>2</sub> ú ERIN <sub>2</sub> e pi is tum a na si ip ri
217	TXM_cuneif_Ha_S_Abb_2_4	sá ka pí im la na tú ú	ma in ne ez bu ú GAN SAHAR i na sa sá ah
218	TXM_cuneif_Ha_S_Abb_2_4	di te+em ka la aš pur am	ma si pi ir l-im sa ih he ru ú la i
219	TXM_cuneif_Ha_S_Abb_2_5	(r) EN ZU li dam ú b	ma um ma em mu re bil me l ka lu

479 occurrences 201-300/479 Nombre de lignes : 100 Exporter

Tablettes cunéiformes en ancien babylonien (Béranger et al., 2015)

qgraal\_cm - 160a

<160a>

Ici commence la version de la *Quête del saint Graal* donnée par le manuscrit K (Bibliothèque Municipale de Lyon, Palais des Arts n° 77), folios 160 recto à 224 verso. Tout le début du texte a été mutilé : la première grande lettre a été découpée, comme on le voit sur la reproduction du manuscrit, et quelques lignes du texte manquent, que nous donnons ici entre crochets, en bleu, d'après le manuscrit Z (Paris, BNF n. acq. fr. 1119, folio 138 recto, colonne a) qui est un manuscrit proche de celui que nous éditons ici.

(p. 1)

[§ 1]

[A la veille de la Pentecoste quant li compaignon de la table ronde furent venu a Kamaalot et il orent oí le servisse et l'en voloit metre les tables a heure de] nonne. l'ors en[tra] [a cheval en la] sale une mout be damoisele, et fu venue si grant oíre que bien le pot l'en veoir, car ses chevaux en fu encore toz suanz, et el descent et vient devant le roi si le salue, et il dist que Diex la benie. « Sire, fet ele, por Dieu dites moi se Lar est ceenz. - Oí voir, fet li rois, veez le la. » Si li mostre, ele va maintenant la ou il est, et li dist : « Lancelot je vidi de par le roi Pellés que vos avec moi venez usqu'en cele forest. » Et li li demande a qui ele est. « Je sui, fait

Fragment du ms. Z (BnF, n.a.fr. 1119, col. 138a) à la place d'un fragment manquant du ms. K

Retour au manuscrit K (Lyon, BM, P.A. 77, col. 160a)

160a / 268

Manuscrit médiéval de la Quête du Graal (Marchello-Nizia & Lavrentiev, 2013)

tdm80j - 2

CHAPITRE I.  
 DANS LEQUEL PHILEAS FOGG ET PASSEPARTOUT S'ACCEPTENT RÉCIPROQUEMENT L'UN COMME MAÎTRE, L'AUTRE COMME DOMESTIQUE

En l'année 1872, la maison portant le numéro 7 de Saville-row, Burlington Gardens - maison dans laquelle Sheridan mourut en 1814 - était habitée par Phileas Fogg, esq., l'un des membres les plus singuliers et les plus remarquables du Reform-Club de Londres, bien qu'il semblât prendre à tâche de ne rien faire qui pût attirer l'attention.

A l'un des plus grands orateurs qui honorent l'Angleterre, succédait donc ce Phileas Fogg, personnage énigmatique, dont on ne savait rien.

DANS LEQUEL PHILEAS FOGG ET PASSEPARTOUT S'ACCEPTENT RÉCIPROQUEMENT, L'UN COMME MAÎTRE, L'AUTRE COMME DOMESTIQUE.

En l'année 1872, la maison portant le numéro 7 de Saville-row, Burlington Gardens, - maison dans laquelle Sheridan mourut en 1814, - était habitée par Phileas Fogg, esq., l'un des membres les plus singuliers et les plus remarquables du Reform-Club de Londres, bien qu'il semblât prendre à tâche de ne rien faire qui pût attirer l'attention.

A l'un des plus grands orateurs qui honorent l'Angleterre, succédait donc ce Phileas Fogg, personnage énigmatique, dont on ne savait rien.

default | facs ... 2 / 219

Fac-simile d'un livre de Jules Verne (Heiden, 2016)

TXM

P158 4 avril 2014 prTXManonymisev2.mp4

file://home/mdecorde/TXM/corpora/p158/HTML/P158/default/P1

P: 027:53 comment

E: 027:54 la chaleur

P: 027:55 alors la chaleur alors ça c' était non 027:57 vous aviez pas fait l' hypothèse de la chaleur pour les raies puisque la cha la température pas la chaleur la température ça dit de quelle couleur globalement on va pouvoir disposer dans le spectre 028:00 les chaudes en haut les plus froides en bas bien et le soleil vous voyez que le soleil en gros est intermédiaire hein puisque dans le soleil il y a toutes les couleurs de l' arc en ciel , dans la lumière émise par le soleil 028:04 bien donc ça c' est les spectres d' émission 028:09 alors j' aimerais aussi que vous repérez et ça serait quand même pas mal maintenant quand vous avez un peu de recul dans quelles activités on a étudié des spectres d' émission 028:43 dans l' activité

E: 028:42 1

P: 028:43 > 028:46 alors dans l' activité 1 oui parce que on a observé bien sur les néons ah les néons ils observent de la lumière voilà vous mettez le néon ici 028:52 fie

REPRENDRE 28:2175 Répéter Taux: Vol: 6 / 12

P158: "lumière" [rppos="V-"]

ref	Contexte gauche	Pivot	Contexte droit
P158 4 avril 2014 prTXManonymisev2, E, 012:18	toutes les zones avancent là bas plus la	lumière est	blanche ben plus il y a de lumière ce spectre là ça
P158 4 avril 2014 prTXManonymisev2, P, 017:55	ou il peut être absorbé par raies une	lumière peut	être absorbé par raies ou par bandes d'accord et en fait
P158 4 avril 2014 prTXManonymisev2, P, 028:09	de l'arc en ciel, dans la	lumière émise	par le soleil bien donc ça c'est les spectres d'émission
P158 4 avril 2014 prTXManonymisev2, P, 033:37	elle aurait un spectre de ce type là	lumière émise	par la tige de fer OK le fer chauffé à blanc c'

Afficher paramètres

Enregistrement audiovisuel d'un cours de physique (projet coord. Par A. Tiberghien, ICAR) (Heiden, 2015)

# Panorama de méthodes textométriques

## TEXTES

Contextes, cotextes  
(locuteur, période...)  
« tenants »



## MOTS

Unités, motifs, traits...  
« contenus »

**Spécificités**

**Retour au texte**

Lecture, édition du  
texte, visualisation du  
document source

**Classification**

appliquée aux textes,  
analyse arborée

**AFC**

analyse factorielle des  
correspondances

**Cooccurrences**

**Vocabulaire**

**Segments répétés**

inventaire distributionnel

**Concordance**

**Méthode ALCESTE**

méthode Reinert

Détection et mesure de l'**irrégularité de répartition**

des mots entre les textes :

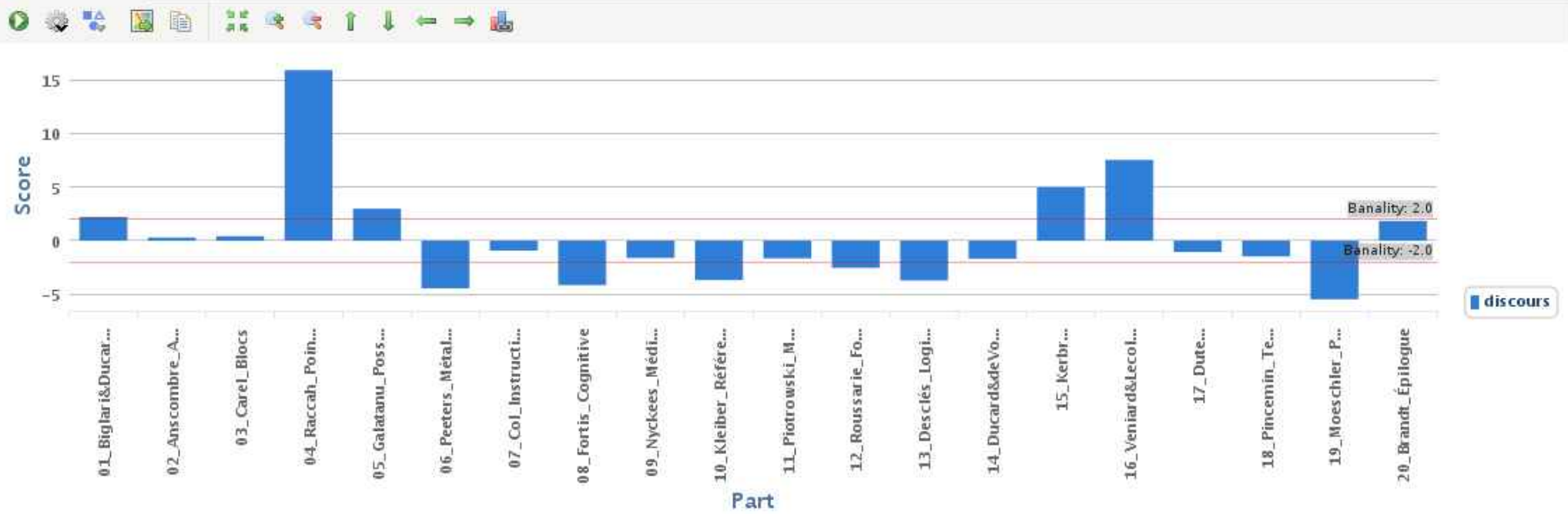
sur-emplois, sous-emplois, liens mots ↔ parties

# Illustration sur le corpus des 20 chapitres de *La sémantique au pluriel*



Property	word													
Units	▲	Frequency	01_BiglariDucar...	Index	02_Anscombe...	Index	03_Carel	Index	04_Racchah	Index	05_Galatanu	Index	06_Peeters...	Index
discontinus		1	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0
disconvenance		1	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0
<b>discours</b>		<b>257</b>	<b>19</b>	<b>2.2</b>	<b>15</b>	<b>0.3</b>	<b>19</b>	<b>0.4</b>	<b>54</b>	<b>16.0</b>	<b>24</b>	<b>3.0</b>	<b>1</b>	<b>-4.5</b>
Discours		1	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0
discourse		1	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0
Discourse		1	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0
discréditer		1	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0	0	-0.0

[discours]



## Spécificités du mot « discours » selon les chapitres

Dans 4 chapitres, le mot est employé significativement plus que dans la moyenne de l'ouvrage. C'est une information quantitative, qui reste à interpréter qualitativement.



# Un autre usage des spécificités : le repérage de (quasi-)exclusivités

- *Anscombe* : **topoï** (52+2 occ. dans le prologue / 55 occ. en tout)
- *Carel* : **bloc** (34+1 / 39 – occ. restantes = 2 renvois + 2 locutions "en bloc")
- *Raccah* : point (105 / 295), vue (99 / 217)
- *Col* : scène (60+1 / 78)
- *Fortis* : prototype (38 / 49)
- *Nyckees* : **opérativité** (31+1 / 32)
- *Desclés* : **combinateur** (32 / 32)
- *Duteil-Mougel* : sème (103 / 120), isotopie (50 / 58), palier (44+3 / 50)
- *Pincemin* : **textométrie** (36+2 / 38), spécificité (20 / 30)
- *Moeschler* : **implicature** (73+3 / 77)

=> des indicateurs possibles ?

- d'emploi terminologique
- de connotation, au sens de l'association sémantique du mot à un contexte théorique, un courant de pensée.

# Panorama de méthodes textométriques

## TEXTES

*Contextes, cotextes  
(locuteur, période...)*  
« tenants »



## MOTS

*Unités, motifs, traits...*  
« contenus »

**Spécificités**

**Retour au texte**

Lecture, édition du  
texte, visualisation du  
document source

**Classification**

appliquée aux textes,  
analyse arborée

**AFC**

analyse factorielle des  
correspondances

**Concordance**

Méthode **ALCESTE**

méthode Reinert

**Cooccurrences**

**Segments répétés**

inventaire distributionnel

**Vocabulaire**

Examen méthodique des **contextes d'emploi**, dans un tableau synthétique permettant des tris et la superposition des répétitions ; herméneutique des passages parallèles

# Illustration sur le corpus des 20 chapitres de *La sémantique au pluriel*



SEMANTIQUES22-230308/<[frlemma="fait"]> [frlemma="fait"]

ref	Left context	Pivot	Right context
p. 369, Duteil-Mougel, Textuelle	l'argumentation et de la succession des «	faits	», ruptures de « points de vue » – correspondent à
p. 317, Kerbrat-Orecchioni, Interactionne	décrit comme suit : (1) A	fait	à B une proposition lexicale : le signifiant « pique-nique » est
p. 217, Kleiber, Référentielle	on nie la réalité – attitude tout à	fait	possible comme en témoigne notre raisonnement – ce ne peut être n'
p. 196, Nyckees, Médiationniste	2016b). Il serait donc tout à	fait	illusoire de prétendre tracer une frontière nette et objective entre la quête
p. 215, Kleiber, Référentielle	projeté nous semble à cet égard tout à	fait	superflue, puisque la précision qu'il ne s'agit que de
p. 318, Kerbrat-Orecchioni, Interactionne	. En voici un autre exemple tout à	fait	similaire, mais emprunté à un article de presse (rapportant un
p. 139, Peeters, Métalangue	par des pronoms non universels mais tout à	fait	courants et plus intelligibles du point de vue de l'apprenant.
p. 129, Peeters, Métalangue	is not something big) paraît tout à	fait	inacceptable. Il y a lieu de croire que [B1]
p. 131, Peeters, Métalangue	mais qui n'en ont pas tout à	fait	le sens. Les primitifs wierzbickiens, par contre, sont lexicalisés
p. 408, Moeschler, Pragmatique	que ce type de sémantique peut tout à	fait	être complété par une pragmatique à la Grice : c'est en
p. 160, Col, Instructionnelle	nouveau spectacle » proposent un sens tout à	fait	perceptible. Cette bonne forme va cependant être modifiée lorsque voilà,
p. 035, Anscombe, Argumentative	énoncés de la première série sont tout à	fait	banals, alors que ceux de la seconde paraissent étranges, et
p. 346, Duteil-Mougel, Textuelle	passage du fait naturel de la consanguinité au	fait	culturel de l'alliance » (ibid.). Une fois
p. 194, Nyckees, Médiationniste	une dimension empirique, nous semble imputable au	fait	qu'elles laissent systématiquement de côté la moitié du problème qui se
p. 088, Raccah, Points-de-vue	possession de celles qui donnent du pouvoir au	fait	d'avoir des amis : le jugement de valeur sur les '
p. 041, Anscombe, Argumentative	représenteraient. Ce phénomène est à relier au	fait	que les phrases génériques habituelles admettent des exceptions sans cesse
p. 077, Raccah, Points-de-vue	on peut fournir une explication sémantique simple au	fait	que, généralement, les énoncés de la phrase : Max est
p. 410, Moeschler, Pragmatique	Q et ensuite P] Cela tient au	fait	que et doit être interprété comme 'et ensuite', mais
p. 128, Peeters, Métalangue	qu'un mot quelconque (ou un autre	fait	de langue) signifie pour ceux qui parlent la langue à laquelle
p. 395, Pincemin, Textométrie	rejet de conventions, etc. Quant aux	faits	singuliers, signaux faibles au plan quantitatif alors que parfois majeurs au
p. 091, Raccah, Points-de-vue	. Ce qui est particulièrement intéressant dans ce	fait	apparemment anodin, c'est que les interlocuteurs se trompent très raremen
p. 108, Galatanu, Possibles	concernée, ancrées culturellement et, de ce	fait	, cinétiques. Soulignons quelques spécificités des stéréotypes en SPA 24.
p. 068, Carel, Blocs	signification de mauvais temps et, de ce	fait	, d'allure naturelle, doxale. (On pourrait bien sûr
p. 234, Piotrowski, Morphodynamique	de morphologies signifiantes et ne constitue de ce	fait	que la machinerie où se trame. alors au premier plan.

1 - 100 / 188

## Concordance du mot « fait » (lemme), triée sur le contexte gauche.

*Repérage de différents cas : locutions, erreur d'étiquetage, ... Dans un premier temps pour l'analyse, on choisira de s'en tenir à la forme plurielle (43 / 188 occ.), qui concentre la majorité des emplois hors locution (28 / 43).*

# Panorama de méthodes textométriques

## TEXTES

*Contextes, cotextes  
(locuteur, période...)*  
« tenants »



## MOTS

*Unités, motifs, traits...*  
« contenus »

**Spécificités**

**Retour au texte**

Lecture, édition du  
texte, visualisation du  
document source

**Classification**

appliquée aux textes,  
analyse arborée

**AFC**

analyse factorielle des  
correspondances

**Cooccurrences**

**Vocabulaire**

**Segments répétés**

inventaire distributionnel

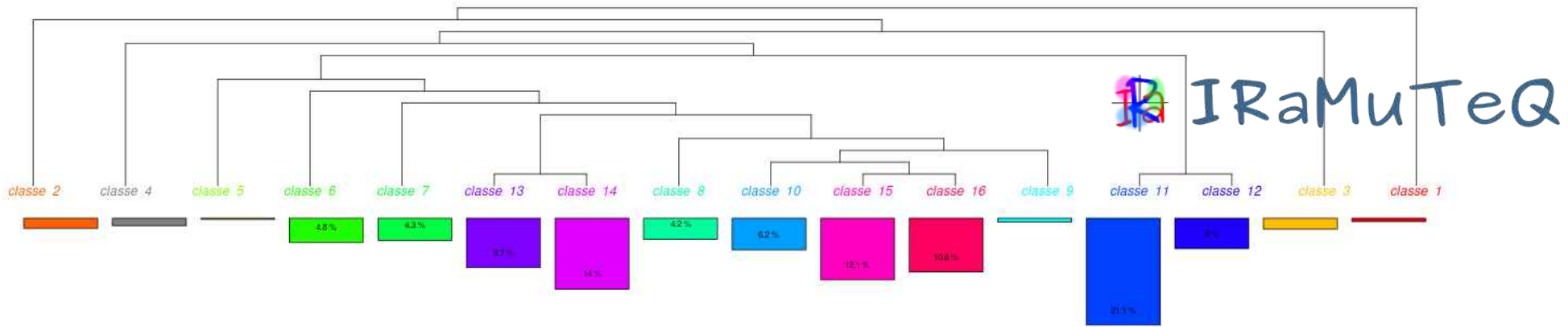
**Concordance**

**Méthode ALCESTE**

méthode Reinert

Induction d'univers sémantiques, de **champs lexicaux**,  
à partir du corpus

# Illustration sur le corpus des 20 chapitres de *La sémantique au pluriel*



pt	implica	savoir	msn	corpus	opérati	formel	spa	ducrot	discour	énoncé	riche	vrai	md	semer	this
neg	implicati	professi	primitif	textomét	cognitio	sémant	discursii	topoi	interprét	unité	ami	entité	signifiar	isotopie	the
concrétis	implicati	déontiqu	wierzbic	statistiqu	expérien	logique	galatanu	anscom	segment	instructic	max	convoq	signifiés	afférent	is
battre	explicatur	épistémi	anglais	textométri	humain	théorie	noyau	tbs	ressourc	construc	raciste	scène	substan	récurrenc	of
guerre	généralis	moral	goddard	calcul	monde	langage	potentialii	locuteur	pratique	sentir	possessi	schème	sémiolin	inhérent	it
fiils	conventio	apprendre	universel	fréquence	historicit	linguistiq	stéréotyp	argument	participa	construire	républicai	fille	signe	ex	at
menaçant	présuppo	acteur	métalang	logiciel	saussure	étude	enseigna	intention	profond	contraint	intelligent	prédicat	fonction	sémèmes	that
aspect	explicatur	élève	peeters	analyse	langage	pragmatiq	sémantic	gradabilit	genre	col	obliger	procès	parcours	palier	time
andromaq	contextual	écouter	explicitatic	textométri	épistémic	année	putnam	argument	compte	imposer	imposer	temporel	stabilisati	1994a	for
émouvant	truc	pragmatiq	explicitatic	répartition	médiator	programm	cinétisme	informatif	mettre	décrire	paradoxal	verbe	différentie	1996a	after
froid	critère	école	français	spécificité	conceptu	classique	axiologiqu	destinait	sd	description	lecteur	dénoter	connexio	isotopies	someone
pyrrhus	contextuel	appartenir	scénario	pincemin	significati	montague	déploieme	topo	dénomina	fournir	voir	instant	rapport	rastier	some
dépenser	suppositio	communau	molécule	découpag	communi	courant	reconstrui	enchaîner	interactio	voir	validité	spatial	contenu	générique	and
avare	im	volonté	version	alceste	pensée	réalité	reconstrui	explicatio	textuels	instruction	facile	ego	opposition	spécifique	be
émousser	abi	société	indéfiniss	avenir	réalité	reporter	filliation	raccah	texte	fur_et_à_r	banal	dormir	légalité	lexies	one
dépensier	mercedes	éducation	fragment	évaluer	réel	philosophi	cozma	orecchioni	usage	fuchs	figurer	aimer	conscienc	sémème	you
dissuader	coûter	acquérir	universalit	lexicométri	scientifiqu	linguiste	discursifs	kerbrat	valeur	assembla	suggérer	talmy	structural	lexie	when
empêcher	conversati	compétenc	publier	cooccurr	métaphor	notammer	argument	repandre	choix	compte	membre	colère	contrôle	sémiques	somethin
émouvoir	présuppos	correction	naturel	concordar	égard	mathémat	significatic	attitude	doute	victorri	solution	événeme	plan	sémisme	can
rat	réussir	vocation	naturel	concordar	individuel	cognitif	argument	argument	locuteurs	rendre	conférer	événe	frontière	sauce	many
malheurel	vériconditi	université	original	affinité	sémioticit	science	institution	argument	oral	idéologiqu	chercher	percevoir	instiuer	microsém	do
converse	annuler	motivation	liste	appui	conceptue	tradition	argument	2001a	croynance	perceptio	mot	luc	dynamiqu	isotopant	by
hector	gazdar	confiance	manuel	lemme	1998b	benveniste	searle	respecter	régularité	ny	juger	énonciate	enseigner	taxème	not
économe	euro	république	inventaire	voeu	histoire	domaine	associatio	soutenir	prendre	présuppos	propos	jean	topologiq	macrosem	with
transgress	wilson	manuel	1999b	tafon	conceptio	interface	explicatif	sentenciel	analyste	capable	interlocute	moment	singularité	rx4	ground
negdépen	étudiant	service	exhaustif	texte	historique	approcher	orientatio	notation	pagé	informatio	argumenta	opérateur	éduquer	corbeau	sentence
astyanax	pré	numérique	pédagogic	manuel	parent	syntaxe	dénotatif	signaler	permettre	contribuer	comport	mouveme	différentie	morphème	he
craindre	pragmatiq	institution	natif	traitement	subjectivit	chapitre	communa	propos	tenir	adéquat	ironique	jeune	ab	scénic	meaning
utile	ic	tableau	explicit	cooccurrer	socio	ouvrage	alliance	communic	sembler	effectuer	lire	figure	substrat	transport	no
blesable	enrichisse	penser	natural	lemmatisat	nvckeas	herméne	2018a	communic	interpréte	schématiq	supporter	posséder			

## Méthode Reinert

Mots (lemmes) caractéristiques de 16 classes de segments de texte de l'ordre de 1 à 3 phrases (~40 mots). Noter qu'un même mot peut apparaître dans plusieurs classes (ex. ici « texte »), ce qui est intéressant du fait des variations de sens (acceptations, polysémie).

# Panorama de méthodes textométriques

## TEXTES

*Contextes, cotextes  
(locuteur, période...)*  
« tenants »



## MOTS

*Unités, motifs, traits...*  
« contenus »

**Spécificités**

**Retour au texte**

Lecture, édition du  
texte, visualisation du  
document source

**Classification**

appliquée aux textes,  
analyse arborée

**AFC**

analyse factorielle des  
correspondances

**Concordance**

**Cooccurrences**

**Segments répétés**

inventaire distributionnel

**Vocabulaire**

**Méthode ALCESTE**

méthode Reinert

Détection et mesure des **attirances entre mots**  
(lexicales, phraséologiques, syntaxiques, thématiques...)

# Illustration sur le corpus des 20 chapitres de *La sémantique au pluriel*



SEMANTIQUES22-230308/<[word="disco... ⌘	SEMANTIQUES22-230308/<[word="discou... ⌘	SEMANTIQUES22-230308/<[word="dis... ⌘	SEMANTIQUES22-230308/<[word="di... ⌘																			
Query                         .div_author="Raccah"	Query                         .div_author="Galatan"	Query                         div_author="Kerb"	Query                         iv_author="Venia"																			
Cooccurrent	Frequ	CoFr	Sc	Mean	dist	Cooccurrent	Frequ	CoFr	Sc	Mean	dist	Cooccurrent	Frequ	CoFr	Sc	Mean	dist					
vue	219	11	8	6.2		analyse	295	10	9	2.1		discours	257	8	6	6.2		circulants	3	2	4	.0
énoncés	199	9	6	2.7		officiel	3	3	8	.7		monologal	2	2	5	.0		Sémantique	22	3	4	1.0
points	102	7	6	7.4		enseignant	18	4	7	6.5		dénommer	4	2	4	3.5		discours	257	6	4	3.3
les	2956	34	5	3.0		investissement	7	2	4	7.5		échantillon	4	2	4	1.0		analyse	295	6	4	2.5
imposent	22	4	5	7.8		des	3568	18	3	3.5		tel	96	4	3	3.0		lexicale	67	3	3	4.3
concernent	8	3	5	2.0		l'	3601	18	3	3.6		analyse	295	6	3	2.8		SD	23	2	2	5.0
M	8	3	5	4.0		univers	23	2	2	2.0		travail	54	3	3	5.7		articulation	24	2	2	4.5
idéologies	9	3	5	2.7		Cet	24	2	2	6.5		opposer	12	2	3	5.0		pris	24	2	2	4.0
ces	432	10	4	2.3		discursifs	28	2	2	4.0		participants	17	2	3	6.5		du	1835	11	2	1.1
mots	273	8	4	4.8		départ	31	2	2	3.0		aussi	298	5	2	3.6		niveaux	44	2	2	5.0
entendus	3	2	4	3.0		discours	257	4	2	5.0		»	1421	11	2	5.4		et	4160	18	2	3.4
subjectif	4	2	3	7.0		celui	123	3	2	4.0		discipline	33	2	2	7.5		dimension	60	2	2	5.0
observation	24	3	3	3.3		second	36	2	2	4.5		quoi	39	2	2	7.0		en	2128	11	2	2.5
que	2132	23	3	4.8		sémantique	786	6	2	4.5		de	8428	36	2	4.0		?	185	3	2	1.0
avez	6	2	3	4.5		référentiel	56	2	2	1.5		objet	139	3	2	2.3						
polyphonique	7	2	3	1.0		construire	59	2	2	8.5		interaction	49	2	2	1.0						
portées	7	2	3	2.0		sur	807	6	2	3.5		voir	161	3	2	4.0						
parlent	9	2	3	.0		-	392	4	2	5.5												
constituent	41	3	3	4.0																		
subjectives	10	2	3	4.0																		
compréhension	42	3	2	2.0																		
Bakhtine	11	2	2	6.5																		
† pivot 54, v cooc 39, † cooc 196, T corpus 217701						† pivot 24, v cooc 18, † cooc 92, T corpus 217701						† pivot 34, v cooc 17, † cooc 95, T corpus 217701										

P.-Y. Raccah

O. Galatanu

C. Kerbrat-Orecchioni

M. Veniard & M. Lecolle

## Cooccurrences du mot « discours » à l'intérieur de quatre chapitres

*Pour ce calcul, le contexte a été défini comme une fenêtre de 10 mots avant et après.*

*Ici, vue la (petite) taille de notre corpus, une simple concordance serait plus fructueuse*

*qu'un calcul statistique qui restitue seulement quelques caractéristiques saillantes.*

# Illustration sur un corpus plus large, de discours présidentiels français

Table 2

Co-occurrences of "work" in the corpus Sarkozy compared to de Gaulle

SARKOZY		DE GAULLE	
Words	Deviations	Words	Deviations
réhabiliter (regenerate)	+8.48	technique	+4.79
fruit	+6.13	rendement (productivity)	+4.34
effort	+5.95	production	+3.93
merit	+5.94	information	+3.90
partage (sharing)	+5.41	capital	+3.60
revalorisation (increase, adjustment)	+5.11	échelle (scale)	+3.50
libérer (to free)	+4.76	personnel	+3.24
possibilité (possibility)	+4.64	déplacement (shift)	+3.23
durée (duration)	+4.63	emploi (job, employment)	+3.22
valeur (value)	+4.61	commission (commission)	+3.21
récompense (reward)	+4.32	responsabilité (responsibility)	+3.19
formation (training)	+4.18	jeune (young)	+3.18
réhabilitation (rehabilitation)	+4.16	société (society)	+2.91
taxer (to tax)	+4.14	professionnel (professional)	+2.87
vivre (to live)	+3.85	intérêt (interest)	+2.84
création (creation)	+3.59	œuvre (piece of work)	+2.82
récompenser (to reward)	+3.49	direction	+2.76

Tiré de :  
**Damon Mayaffre.**  
 « Quantitative Linguistics and Political History ». in Jacqueline Léon; Sylvain Loiseau. *History of Quantitative Linguistics in France*, 24, RAM - Verlag, pp.94-119, **2016**, Studies in Quantitative Linguistics, 978-3-942303-48-4. (hal-01454671)



## Cooccurrences du mot « travail » dans les discours de Sarkozy et de De Gaulle

En opérant sur plusieurs centaines de contextes, le calcul statistique synthétise l'environnement lexical d'un même mot, différent dans la bouche de deux locuteurs.



# Panorama de méthodes textométriques

## TEXTES

*Contextes, cotextes  
(locuteur, période...)*  
« contenant »



## MOTS

*Unités, motifs, traits...*  
« contenus »

**Spécificités**

**Retour au texte**

Lecture, édition du  
texte, visualisation du  
document source

**Classification**

appliquée aux textes,  
analyse arborée

**AFC**

analyse factorielle des  
correspondances

**Concordance**

**Cooccurrences**

**Vocabulaire**

**Segments répétés**

inventaire distributionnel

**Méthode ALCESTE**

méthode Reinert

Détection et fréquence des **enchaînements de mots récurrents**  
(mots composés, figement, phraséologie...)

# Illustration sur le corpus des 20 chapitres de *La sémantique au pluriel*



Corpus SEMANTIQUES22-230308

SEMANTIQUES22-230308 / <[...]

Properties SEMANTIQUES22-230308 / <[...]

Query [word!="\p{P}"]{3,7} Properties: word Edit

word	Frequency
En d' autres termes	21
Desclés et al	16
d' une part	15
d' autre part	11
Dans cette perspective	9
à l' inverse	8
pour ainsi dire	8
émouvant DC Y est ému par X	7
est émouvant DC Y est ému par	7
et d' autre part	7
Goddard et Wierzbicka	7
La notion de	7
X est émouvant DC Y est ému	7
ce que nous croyons être la réalité	6
d' un côté	6
la construction du sens d' un énoncé	6

Macro New macro all

cqp

- CQL2Partition
- CQL2SubCorpus
- CompressCQPIndexes
- CreateCQPListFromFile
- CreateCQPListFromGroov
- CreateCQPListFromString
- ExecCQP
- ListMatches
- Positions2SubCorpus
- SetMatchingStrategy**
- ShrinkCQLValues

Console System output

Index of <[word!="\p{P}"]{3,7}>, property @word, in SEMANTIQUES22-230308 corpus...  
78.684 item for 80.152 occurrences.

## Recherche de séquences de mots répétées

Ponctuation exclue, longueur entre 3 et 7 mots. On recueille des locutions, mais aussi certaines « briques expressives » récurrentes.

# Panorama de méthodes textométriques

## TEXTES

*Contextes, cotextes  
(locuteur, période...)*  
« tenants »



## MOTS

*Unités, motifs, traits...*  
« contenus »

**Spécificités**

**Retour au texte**

Lecture, édition du  
texte, visualisation du  
document source

**Classification**

appliquée aux textes,  
analyse arborée

**AFC**

analyse factorielle des  
correspondances

**Cooccurrences**

**Concordance**

**Méthode ALCESTE**

méthode Reinert

**Vocabulaire**

**Segments répétés**

inventaire distributionnel

Inventaire des **mots avec leur fréquence d'emploi**  
dans le corpus : attestations, dominantes lexicales

# Illustration sur le corpus des 20 chapitres de *La sémantique au pluriel*



SEMANTIQUES22-230308/<[frpos=...

Query Properties: word Edit

word	Frequency
sens	998
p	408
sémantique	396
signification	380
exemple	345
langue	329
unités	288
analyse	274
mots	273
discours	256
énoncé	228
vue	215
théorie	211
mot	202
énoncés	196
forme	196
point	191
cas	187
langage	181
langues	178
monde	175

1 t 43816, v 5616, fmin 1, fmax 998

*Noms communs  
(formes graphiques)*

SEMANTIQUES22-230308/<[frpo...

Query Properties: frlemma

frlemma	Frequency
sen sens	1001
signification	532
langue	515
mot	478
unité	458
sémantique	455
énoncé	424
p	408
exemple	406
analyse	325
forme	322
théorie	312
point	295
discours	257
terme	251
relation	232
locuteur	227
principe	225
texte	221
vue	217
type	216

1 t 43816, v 4358, fmin 1, fmax 1001

*Noms communs  
(lemmes)*

SEMANTIQUES22-230308/<[frpos=...

Query Properties: frlemma E

frlemma	Frequency
sémantique	664
autre	587
même	461
linguistique	406
différent	228
lexical	190
discursif	170
pragmatique	142
cognitif	134
nouveau	120
formel	115
possible	114
grand	110
général	104
référentiel	96
particulier	93
verbal	93
simple	90
français	89
complexe	88
théorique	88

1 t 15912, v 2259, fmin 1, fmax 664

*Adjectifs qualificatifs  
(lemmes)*

SEMANTIQUES22-230308/<[frpo...

Query Properties: frlemma

frlemma	Frequency
être	4587
avoir	1474
pouvoir	918
faire	547
dire	418
voir	367
donner	280
permettre	271
devoir	196
prendre	191
trouver	188
agir	181
mettre	180
savoir	170
vouloir	167
constituer	165
partir	164
comprendre	163
rendre	163
définir	162
construire	152

t 26098, v 1573, fmin 1, fmax 4587

*Verbes  
(lemmes)*

**Inventaires fréquentiels de mots**, ici par catégorie grammaticale.

On peut les lire triés par fréquence (dominances) ou alphabétiquement (dictionnaire minimal du corpus). Des « intrus » sont dus à des erreurs de l'analyseur morphosyntaxique.

# Principes opératoires mis à l'épreuve

Une sémantique...

1. Textuelle -endogène, relative
2. Distributionnelle
3. Différentielle
4. Matérielle
5. Interprétative
6. Ouverte

# Principes opératoires mis à l'épreuve

## 1. Textuelle -endogène, relative

- Articulation global (intertexte, corpus) / local (lexique, morphosyntaxe) : les équilibres du corpus servent de référence (modèle, jauge) ;
- Le corpus est déterminant, et seul nécessaire (apport facultatif d'informations externes)

## 2. Distributionnelle

## 3. Différentielle

## 4. Matérielle

## 5. Interprétative

## 6. Ouverte

# Principes opératoires mis à l'épreuve

1. Textuelle -endogène, relative
2. Distributionnelle
  - Contextes de tous ordres :  
syntagme, paragraphe, texte,  
genre textuel ou autres groupements de textes...
3. Différentielle
4. Matérielle
5. Interprétative
6. Ouverte

# Principes opératoires mis à l'épreuve

1. Textuelle -endogène, relative
2. Distributionnelle
3. Différentielle
  - Analyse comparative, contrastive
  - Éventuellement référentielle et situationnelle par l'annotation, les métadonnées
4. Matérielle
5. Interprétative
6. Ouverte



# Principes opératoires mis à l'épreuve

1. Textuelle -endogène, relative
2. Distributionnelle
3. Différentielle
4. Matérielle
  - Données factuelles : observables, mesures
  - Explicitation (formalisation) des traitements et de leurs paramètres
5. Interprétative
6. Ouverte

# Principes opératoires mis à l'épreuve

1. Textuelle -endogène, relative
2. Distributionnelle
3. Différentielle
4. Matérielle
5. Interprétative
  - Parcours d'analyse et pluralité des lectures
  - Implication du chercheur, contrôle de l'interprétation
  - Évaluation par le cheminement plus que par le résultat
6. Ouverte

# Principes opératoires mis à l'épreuve

1. Textuelle -endogène, relative
2. Distributionnelle
3. Différentielle
4. Matérielle
5. Interprétative
6. Ouverte
  - Analyse jamais complète ni définitive : approche par convergence, découverte par exploration
  - Vertus heuristiques de l'automatisation des traitements

# Un mot sur la question des unités linguistiques pour l'analyse sémantique

- La textométrie travaille-t-elle sur des unités vraiment linguistiques ?
  - Des mots, ou un découpage technique de chaînes de caractères ?

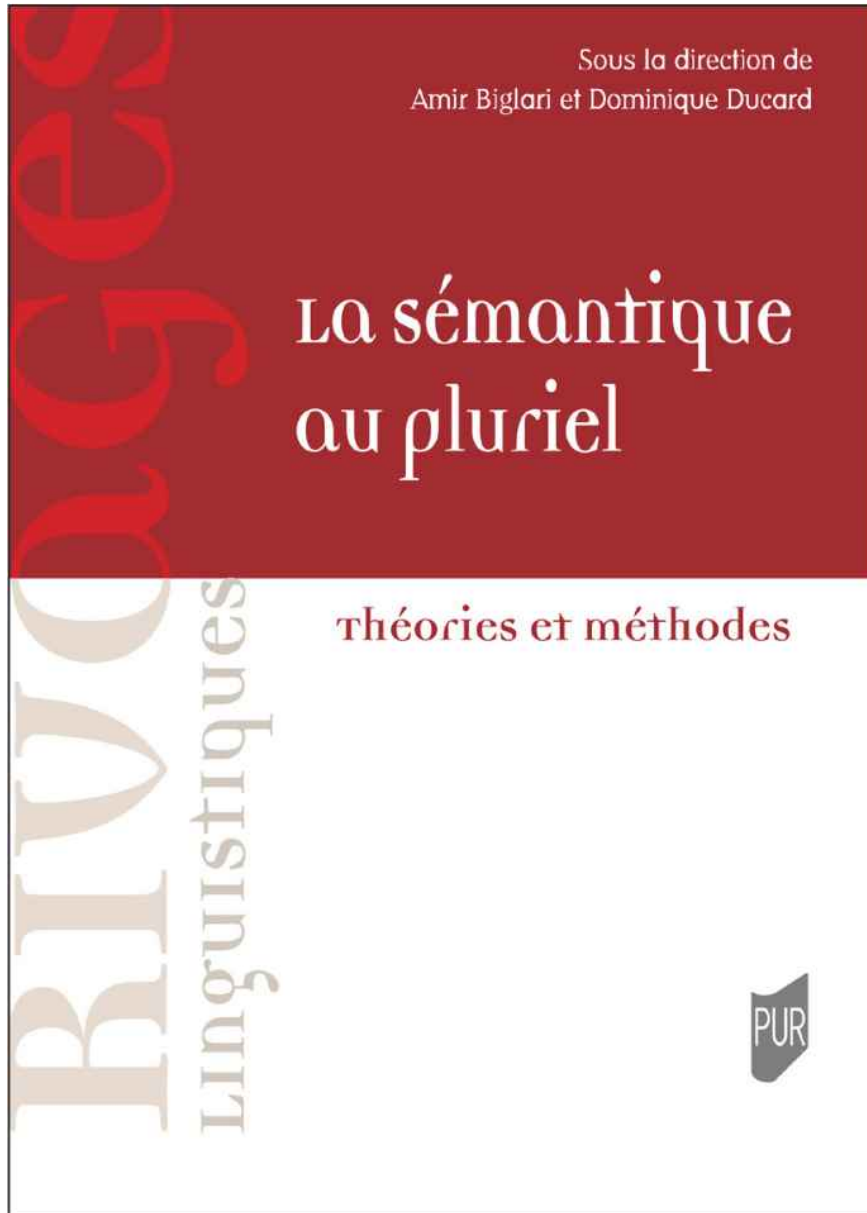
# Un mot sur la question des unités linguistiques pour l'analyse sémantique

- Techniquement, la méthode utilise deux informations :
  - l'occurrence (ou *token*) (délimitation)
  - le type (identification).
- Mais types et occurrences ne sont pas imposés par la méthode !
  - responsabilité du chercheur
  - par exemple :
    - Préciser la façon de découper les mots, donner les frontières de tout ou partie des mots du corpus
    - Choisir de travailler sur les formes graphiques, les lemmes, combiner différentes informations lexicales et morphosyntaxiques, annoter et utiliser des catégories définies et adaptées à la recherche...

# Un mot sur la question des unités linguistiques pour l'analyse sémantique

- Articulation entre
  - unités élémentaires
    - au départ, pour donner prise à l'analyse
    - pragmatiques, (très) approximatives
    - robustesse → capacité à travailler sur des données « réelles », imparfaites et volumineuses
  - et
  - unités descriptives
    - pas données mais construites au fil de l'analyse ; contextuelles, voire dynamiques.
- Autrement dit, le chercheur n'est pas « dupe » des moyens initialement frustes mis en œuvre

# Pour poursuivre la réflexion...



...comme en bonne  
méthode textométrique,  
le retour au texte !