

Formation de l'URFIST de Strasbourg, 28 mai 2015

L.PAIRET Université de Picardie

Introduction à la statistique textuelle

- 1 Types de corpus
- 2 Unités de Statistique textuelle
- 3 Concepts de lexicométrie
- 4 Méthodes de visualisation
- 5 Méthodologie avec une étude

Analyse Statistique de textes

- Lors du traitement statistique des réponses à questions ouvertes, la forme des réponses, les connotations véhiculées par des termes apparemment équivalent constituent ... une information très riche qui permet souvent de nuancer ou d'infléchir les contenus plus manifestes que l'on peut repérer.

L.LEBART, A.SALEM

OBJECTIFS

DEPOUILLEMENT DE QUESTIONNAIRES

RECHERCHE DOCUMENTAIRE

TEXTMINING

TYPOLOGIE

Extraire des régularités

ASSOCIATION LEXICALE

Carte factorielle

1 Type de corpus

- Type
- Encodage
- Variables signalétiques

Type de Données Textuelles

- Exploitation des réponses aux questions ouvertes
- Les entretiens, blogs, forums, sites peuvent être assimilés à des réponses à des questions ouvertes de longueurs variables
- Ils sont accompagnés d'informations complémentaires : locuteur, URL, ...

ASSOCIATION

Du texte ?

Fichier de DONNEES TEXTUELLES

IMPORTATION de DONNEES TEXTUELLES

délimiter les observations

exemples: entretiens, discours, blogs ...
enregistrés avec un traitement de texte

sauvegarder au format TEXTE

Détourner

URL	SITE	URLRELATIVE	DATEMODIFICATION	TEXTEINTEGRAL
http://www.cndp.fr/crdp-dijon/Un-professeur-en-ZEP.html	www.cndp.fr			<pre>me="lienmenu"> <div id="menu-haut"> <a "spip.php?rubrique57" title="Re r/" target="blank" alt="Lien vers le site du minist&egrave;re de l'&Ea ation">Documents, actes et rapports pour l&#8217;éducation &nbsp;>&nbs nu-rub">Trois regards sur la responsabilité</s ponsabilite-dans-1.html" title="">La responsabilité dans l'enseigne ZEP, IUFM d'Amiens <div class="doc1"> <a "IMG/pdf/05Responsabilit ne me paraît pas bien convenir, car lorsqu'on est mobilisé par l'espr sme, le sens d'un métier qu'on ne peut pratiquer de manière aseptisée llement (mais pas seulement, j'y reviendrai), afin qu'il ait les atout e pour que ceux-là tirent vraiment parti de l'école. Ma responsabilit se livrer à des exercices d'entraînement...). Il est de notre responsabi agement, ce qui renvoie à la responsabilité... de l'institution. Deuxi émoire des élèves, de ce qu'ils se sont réellement approprié&nbsp;! N sans doute, mais elle traverse tous les niveaux de l'enseignement. Il l;&nbsp;&nbsp;ponts&nbsp;&nbsp;&#187; avec la grande culture&nbsp;&nbsp;'#nb4' class='sp ce qui est plus difficile - de l'entretenir. Jouer sur la valorisation daptées.</p> <h2 class="spip">J'assume</h2> <p>Une question se pose s&nbsp;&nbsp;&#187; s'occupaient bien d'éducation et Jules Ferry n'a jamais iculier jeunes, qui se sont mobilisés autour d'une &#171;&nbsp;&nbsp;certains ire, dans l'intérêt bien compris de tout le monde. Dans un collège ZEP ... ' id='nh6'>6]&nbsp;&nbsp;! Mais je m'insurge aussi contre les collègues qui s de montrer mon respect inconditionnel pour les personnes, au-delà de</pre>

Une base ?

Fichier de DONNEES NUMERIQUES
QUALITATIVES ET TEXTUELLES

IMPORTATION de DONNEES
AUTOMATIQUE

STRUCTURE : FONCTION DU LOGICIEL

Tableur

	1. Variable1	3. TYPE	4. TYPE_1	7. DATE	8. Variable8
1	1	INTERVIEW	INTERVIEW	10/05/1984	"Le scepticisme ambiant sur les chances de réussite
2	2	ALLOCU...	ALLOCU...	05/07/1984	"Monsieur le Maire, Mesdames et Messieurs. Je
3	7	ALLOCU...	ALLOCU...	05/07/1984	Monsieur le Maire, Mesdames et Messieurs, Ainsi
4	8	ALLOCU...	ALLOCU...	06/07/1984	"Messieurs, Messieurs les Présidents, Mesdames et
5	3	ALLOCU...	ALLOCU...	22/11/1986	"Monsieur le Maire, Mesdames et Messieurs les
6	4	ALLOCU...	ALLOCU...	26/11/1986	"Messieurs les Ministres, Mesdames et Messieurs, Je
7	5	INTERVIEW	INTERVIEW	04/12/1987	"Rendez-vous, j'ai toujours pris le temps de
8	6	MESSAGE	MESSAG...	13/12/1987	"Monsieur le Maire, Cher Pierre MAUROY,
9	9	INTERVIEW	INTERVIEW	10/12/1989	"Le rythme a été rapide dans certains pays ; il
10	10	RENCON...	MESSAG...	27/01/1990	"Bonjour, Je vous remercie de votre accueil. J'ai été
11	11	DECLAR...	MESSAG...	17/12/1993	"Je vais vous préciser certaines choses sur ce qu'il
12	12	ALLOCU...	ALLOCU...	31/12/1993	"Mes Chers Compatriotes, 1993 s'en va. Elle aura été
13	13	ALLOCU...	ALLOCU...	04/01/1994	"Monsieur le Vice-Président, Mesdames,

Encodage spécifique pour le traitement

laiciteLEXICO.txt - Bloc-notes

Fichier Edition Format Affichage ?

si on devait pousser à l'extrême ce type de situation toutes les lois difficiles pourraient être arraché qu'on soit pour ou qu'on soit contre le contenu de cette réforme, c'est l'exemple typique de ce qu'il ne rien, personne ne sait rien. Il faudrait pouvoir se reporter au texte qui a été adopté mardi. Je m'y attache c'est aux municipalités de juger, pas au maire, d'ailleurs je ne voudrais pas être à leur place, parce que Ah, cela sera de toute manière écrit dans les textes. Ne vous inquiétez pas !

Je trouve que la laïcité (que de fois l'ai-je entendu dire chez des dirigeants des organisations d'enseignants) Enfin, voilà, je vous remercie de votre démarche, vous avez bien voulu saisir l'occasion de mon passage

<dis=ALLOCUTION><med=POPULATION><lieu=Paris><dat=vendredi31décembre1993><an=93>

Mes Chers Compatriotes, 1993 s'en va. Elle aura été l'année noire du chômage, qui ronge nos sociétés occidentales. Elle aura été l'année où les inégalités grandissantes ont conduit les peuples de l'Est qui l'avaient si bien connue. Elle aura vu l'ancienne Europe soviétique continuer de se disloquer, tandis que la guerre, la misère, et la mort Mais 1993 c'est aussi autre chose que ce tableau tragique, c'est la paix revenue au Cambodge, la fin de la guerre, c'est encore, pour m'en tenir à quelques exemples de ce que peut notre pays, l'admirable découvert de la pyramide de Saï. Comme j'aimerais que tant d'efforts soient récompensés en 1994 dans des domaines tels que l'emploi, l'ordre, la justice. Mes Chers Compatriotes,

Au mois de mars dernier, vous avez élu à l'Assemblée Nationale une nouvelle majorité politique. Après plus de cent ans, la France a derrière elle une longue et noble Histoire. Depuis la Révolution de 89 et la Déclaration des Droits de l'Homme de tous côtés les périls nous pressent. L'Algérie déchirée la Russie incertaine, les Balkans en feu. Et l'Europe. Faut-il craindre le pire ? Non. Nous avons la chance d'être un pays écouté parmi les Nations. Nous avons la chance de voir que le voeu que j'exprime, mes Chers Compatriotes, s'ajoute à ceux que je formule pour vous. J'adresse une pensée particulière à nos Chers Compatriotes, je vous le demande, soyez unis et solidaires. C'est la réponse à tout. Bonne et heureuse Année 1994.

Vive la République !

Vive la France.

<dis=ALLOCUTION><med=CORPS><lieu=PARIS><dat=4JANVIER1994><an=94>

Monsieur le Vice-Président, Mesdames, Messieurs, Nous voici de nouveau pour la plupart d'entre vous et pour moi-même. Je vous remercie Monsieur le Vice-Président des propos que vous venez de tenir, des voeux que vous avez formulés. Le Vice-Président un monde en mutation où se sont atténués des antagonismes dont on ne voyait pas la fin, dont nous sommes assez loin, on en conviendra entre nous, des perspectives que nous faisons voir les tenants et les aboutissants. Ce n'est pas le seul aspect, il n'est pas négligeable, de l'urgence de la construction européenne. Elle est la condition de ce qu'il est convenu d'appeler cette période d'incertitude où nous sommes entrés exige avant tout la cohésion. Vous vous interrogez, Monsieur le Vice-Président, sur les rapports entre "l'esprit démocratique" et "l'efficacité". C'est sur cet ensemble équilibré de fonctions complémentaires qu'a été édifiée la République française. La République doit avoir les moyens de ses missions ; évitons de les dilapider ou de priver l'Etat des ressources primordiales.

Plus généralement, l'Etat républicain ne peut pas renoncer à ses responsabilités sans compromettre l'équilibre.

Tableur ou texte encode

1	INTERVIEW	PRESSE	Paris	Jeudi 10 mai 1984	Le scepticisme ...
2	ALLOCUTION	CORPS	Le PUY	Jeudi 5 juillet 1984	Monsieur le ..

<dis=INTERVIEW><med=PRESSE><lieu=Paris><dat=Jeudi10mai1984>
Le scepticisme ...

<dis=ALLOCUTION><med=COLLECTIVITETER><lieu=LEPUY><dat=Jeudi5juillet1984>
Monsieur le Maire, Mesdames et Messieurs,Je commencer ...

Signalétique

CLÉS ou VARIABLES

MODALITE

<dis=INTERVIEW><med=PRESSE><dat=Jeudi10mai1984>

Le scepticisme ambiant sur les chances de réussite de toute politique économique, je dis bien de toute politique, dure depuis bientôt dix ans. Il a atteint, puis élimine la majorité précédente. Il frappe l'actuelle majorité. A cela une explication : l'opinion

<dis=ALLOCUTION><med=CORPS><lieu=PARIS><dat=4JANVIER1994>

Monsieur le Vice-Président, Mesdames, Messieurs, Nous voici de nouveau pour la plupart d'entre vous et pour moi-même, dans cette salle des fêtes du Palais de l'Elysée pour la cérémonie traditionnelle des vœux, traditionnelle avec ce qu'elle comporte par définition de convenu mais quand même rencontre

NOM de VARIABLE

2 Unité statistique

- Les **unités statistiques** sont découpées dans le texte:
l'unité de base est la *forme graphique*
- L'étude statistique porte sur les
 - fréquences et gamme de fréquences
 - spécificités
 - profils de fréquences des formes graphiques
 - spécificités

Occurrences ou fréquences

Fréquences

$$T = 14 \quad V = 9$$

Le texte
le livre
l'œuvre

Le livre
le titre
la page
le texte

$$F_{\max} = 3 \quad V_1 = 5 \quad V_2 = 3 \quad V_3 = 1$$

$$\sum_{i=1}^{F_{\max}} V_i = V$$

Gammes de fréquences

fréquence	1	2	3
effectif	5	3	1

$$\sum_{i=1}^{F_{\max}} V_i \times i = 14 = T$$

Index

Lexicométrie

Fréquences décroissantes

3	le
2	Le
2	livre
2	texte
1	l'
1	la
1	oeuvre
1	page

Fréquences **MODALES**

Lexicographique

alphabétique

banales

modales

Loi de ZIPF

$n_i \times \text{rang } i = C^{\text{te}}$

Dictionnaires

L'ensemble des formes et segments compose le dictionnaire du corpus

Mais

seuls les mots SIGNIFIANTS formeront le vocabulaire étudié

- "non-dictionnaires "
- dictionnaires spécifiques

Réduction

- Les mots-outils comme les articles par exemple, sont en principe éliminés
- Les unités *lemmatisées* permettent de limiter le nombre de mots-clé (ramener les verbes à l 'infinitif...);
- une lemmatisation peut être complémentaire d'une analyse

Cartographie par réseau

Trier l'information par **mots clés, rechercher les cooccurents**

Les mots clés s'organisent selon un réseau d'**associations, des indices de cooccurrence**

Aide à la détection de **liens** sous-jacents

Représentation suggestive par carte de type taxinomique ou **sémantique**

3 Concepts lexicométriques et sémantique

■ Segments répétés

mots contigus trouvés à plusieurs reprises ensemble

■ Concordance

ensemble des contextes d'un mot

■ Contexte

les mots trouvés à proximité (paramétrage)

■ Cooccurents

les mots trouvés, associés, à plusieurs reprises

Concordance

Ensemble des contextes d'une forme.

Le texte le **livre** l'œuvre

Le **livre** le titre

la page le texte

autant de lignes de contextes que de fréquences:
synthèse

Cooccurrences

Fréquences

T= 14 V= 9 S=7 VS=5

Le texte
le livre
l'œuvre

Le livre
le titre
la page
le texte

ex: Cooccurrences SIGNIFIANTS
BALAYAGE 3 formes

	livre	oeuvre	page	texte	titre
livre	2	2	0	1	1
oeuvre	2	1	0	0	0
page	0	0	1	1	1
texte	1	0	1	2	0
titre	1	0	1	0	1

Cooccurrences modales

Thème principal ou classe:
(œuvre, livre)

Cooccurrences: Pôle

- **Forme-pôle k**
implique **UNE UNITE de CONTEXTE**
balayée autour de k
→ deux sous-parties: avant k après
cofréquence / distance moyenne (nombre
occurrences)
→ **PROBABILITES**
ensemble de formes situées dans le contexte

Cooccurrences multiples: Implication réciproque

- K formes

$$\frac{N_{123\dots k}}{N_1 N_2 \dots N_k}$$

$N_{123\dots k}$ nombre de cooccurrences des k formes dans le corpus

$N_1 N_2 \dots N_k$ nombre d'occurrences de chacune



utilisé en RD indexation automatique

Coefficient d'implication réciproque

	livre	oeuvre	page	texte	titre
livre	2	2	0	1	1
oeuvre	2	1	0	0	0
page	0	0	1	1	1
texte	1	0	1	2	0
titre	1	0	1	0	1

2 formes: {livre oeuvre}, $N_{12}=2$ $N_{11}=2$ $N_{21}=1$ CIR= 1

{livre titre}, $N_{12}=1$ $N_{11}=2$ $N_{21}=1$ CIR= 0,5 {livre page} CIR= 0

SPECIFICITES et parties

FORMES SOUS-REPRESENTEES ou SUR-REPRESENTEES?

n_{ij}	D1	D2
le	1	2
Le	1	1
texte	1	1
livre	1	1

- 3 le
- 2 Le livre texte
- 1 l' la œuvre

	D1	D _j	D _n	
Fi	n_{i1}	n_{ij}	n_{in}	$n_{i.}$
$N_{.1} = 6$				$N_{.2} = 8$
$n_{oeuvre1} = 1$				$n_{oeuvre2} = 0$
$n_{le1} = 1$				$n_{le2} = 0$
				$n_{.j}$
				$n_{oeuvre.} = 3$
				$n_{le.} = 1$
				25
				T

SPECIFICITES

63042 FORME

libertés

14 OBS

Distribution théorique LOI
HYPERGEOMETRIQUE

TYPE DISC

ALLOCUTION

N.1

25700

10

0,01561466

INTERVIEW

N.2

22157

3

0,13489703

RENCONTRE

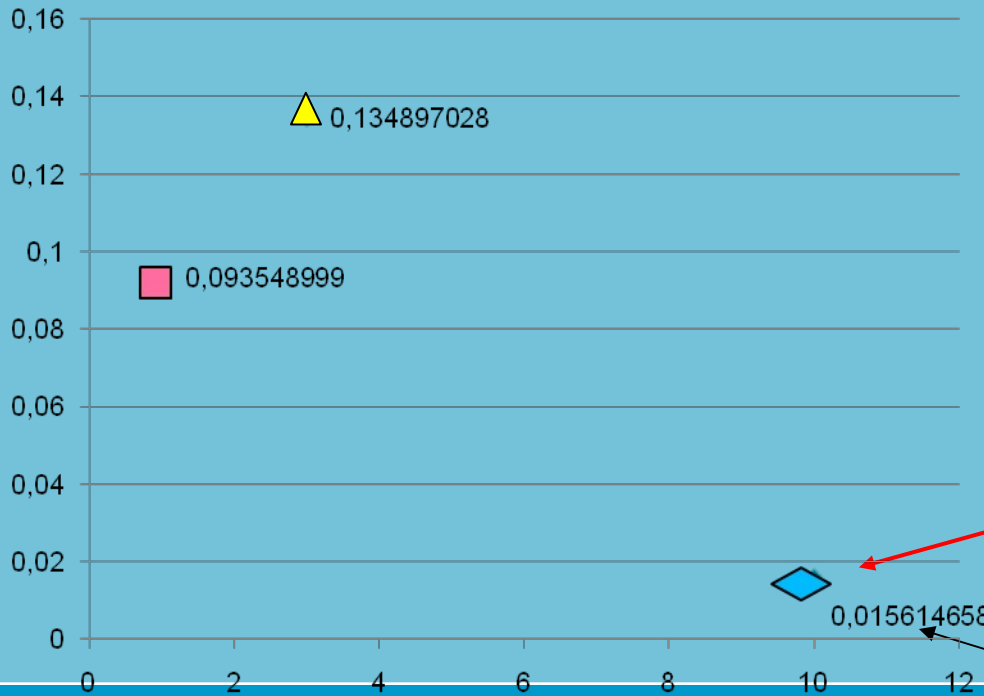
N.3

15195

1

0,093549

◆ Pr théo
d'obtenir
valeur
modale sup à
l'Obs



La somme des
proba. d'avoir
une fréquence
supérieure ou
égale à 10 est
très faible

Mesure de spécificité

Situer f_{ij} par rapport à tous les échantillons $f_{.j}$

$$f_{i/j} = f_{ij} / f_{.j}$$

Sous réserve que l'hypothèse d'indépendance entre les formes soit respectée, utilisation d'une valeur-test

i.e. les parties du corpus sont parfaitement homogènes

estimation de la proportion de formes i
(attendu par la loi hypergéométrique)

dans le corpus,

comparaison avec celle de la partie

Caractéristique

Une forme est **sous-représentée** dans une partie si:

Sa fréquence observée est **inférieure** à la valeur modale attendue et

Si la probabilité d'avoir cette valeur observée est faible

Une forme i sera **caractéristique ou sur-représentée**

Si sa fréquence observée dans la partie j est **supérieure** à la valeur modale attendue et si la probabilité d'avoir cette valeur observée est faible

4 Analyse de Données Textuelles

L'analyse repose sur trois étapes:

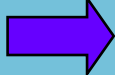
Formes modales

**Formes et segments
caractéristiques**

Analyse des correspondances

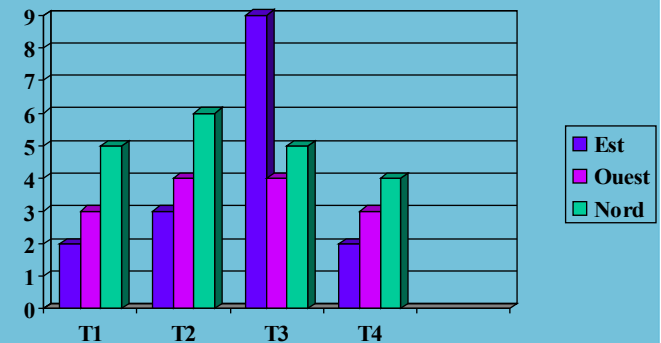
tableaux lexicaux croisement :
variables qualitatives avec les
formes et segments
associations

carte factorielle



Est 16
Nord 15
Ouest 11
Sud 04
Lille 03
Dax 03
Lyon 01

...



Profils lexicaux

Les p mots retenus en ligne,

les n modalités de la variable en colonne



	T1	T2	T3	T4
Est	2	3	9	2
Nord	5	6	5	4
Ouest	3	4	4	3

L'analyse des proximités s'effectue comme en AFC

Distance

Entre points colonnes du corpus

$$d^2(j, j') = \sum_i (1/f_{i.}) (f_{ij}/f_{.j} - f_{ij'}/f_{.j'})^2$$

	i				
f _{ij}	T1	T2	T3	T4	<i>profils lignes</i>
Est	0,04	0,06	0,18	0,04	0,32
Nord	0,1	0,12	0,1	0,08	0,4
Ouest	0,06	0,08	0,08	0,06	0,28
<i>profils col</i>	0,2	0,26	0,36	0,18	

Entre termes i, i' dans le corpus

$$d^2(i, i') = \sum_j (1/f_{.j}) (f_{ij}/f_{i.} - f_{i'j}/f_{i'.})^2$$

Distance entre formes

$$d^2(i, i') = \sum (1/f_{.j}) (f_{ij}/f_{i.} - f_{i'j}/f_{i'.})^2$$

j	i=EST	i'=NORD	
T1	(+1/0,2)	((0,04/0,32)-(0,1/0,4)) ²	0,08
T2	(+1/0,26)	((0,06/0,32)-(0,12/0,4)) ²	0,05
T3	(+1/0,36)	((0,18/0,32)-(0,10/0,4)) ²	0,27
T4	(+1/0,18)	((0,04/0,32)-(0,08/0,4)) ²	0,03
SOMME			0,43

5 Une méthodologie

- 1 Corpus primaire, dictionnaire
- 2 Mise en évidence formes significantes
- 3 La concordance de ces formes significantes
- 4 Calcul du coefficient d'implication réciproque
- 5 Pourcentage de ces formes significantes dans le corpus
- 6 Spécificités
- 7 AFC, CHA
- 8 Reprise du processus avec le corpus réduit

corpus réduit:

fonction de la taille initiale

formes présentant une fréquence fixée,

lemmatisé offre à certaines formes la fréquence nécessaire pour être retenue.

Descriptive :

2756

de

538

ce

206

même

114

elle

MOT

Lexique

538

la

523

vous

197

cette

112

On

INDEX

1494

l

520

ne

192

C

112

leur

LEXICOMETRIQUE

1186

à

483

une

187

ont

111

sans

Lexical

1151

le

453

pour

181

sur

109

aussi

V 7000 T 63000

MAXFR 2700

FR SIGN 179

1128

que

444

n

179

France

106

pays

70

L

1077

est

421

dans

171

être

105

beaucoup

1%V

69

va

1040

et

416

on

169

avec

101

suis

68

depuis

958

les

394

du

167

Et

97

ceux

68

problème

885

qui

378

y

163

aux

106

pays

67

Cela

864

d

372

c

160

Mais

105

beaucoup

67

non

780

pas

318

Je

159

fait

101

suis

66

droit

770

des

285

s

158

comme

94

temps

64

pense

742

en

156

ou

34

88

La

63

encore

729

il

88

moi

87

lui

86

Vous

85

peu

85

son

84

autres

Concordance

Contextes

Segments répétés

DESAMBIGUISATION

dans les lois créatrices d'espaces de **liberté**, celles de Badinter, d'Auroux et de ..
vie quotidienne qu'ils y gagnent en **liberté**, en responsabilité, en emploi et en partage
des lors qu'on leur ajoute ceux de **liberté**, responsabilité, plan, secteur public
de toutes les libertés et donc de la **liberté** d'enseigner. Mais je ne puis, ni ne prise peut-
elle devenir "créatrice de **liberté**" ? Grande question. L'entreprise a toujours soucier
Il n'est pas de **liberté** sans organisation de cette liberté. La as de liberté sans organisation
de cette **liberté**.

La concurrence à l'état pur, c'est nature de la campagne menée au nom de la **liberté** de l'
enseignement a pu troubler certains

murs de nos villes et représentant la **liberté** brisant l'un des barreaux derrière lesquels le –
là qui exaspèrent les passions. La **liberté** a partout gagné du terrain depuis trois

et sans l'initiative, et sans la **liberté**. Quelle leçon allons-nous nous adresser

le premier prix au service de la **liberté**? Je suis prêt à partir d'ici ex – aequo

comme un abus. On confond souvent **liberté** et privilèges et je demande que l'examen,

pour savoir où commence et où finit la **liberté**, là où commence le privilège. Quand on

c'était cela le respect de la **liberté** Vous croyez normal qu'il y ait un établissement v

vois pas pourquoi je serais privé de la **liberté** que vous venez de nous accorder, M. le

où se trouvent, au Puy, les amis de la **liberté** ..

Mesure d'association

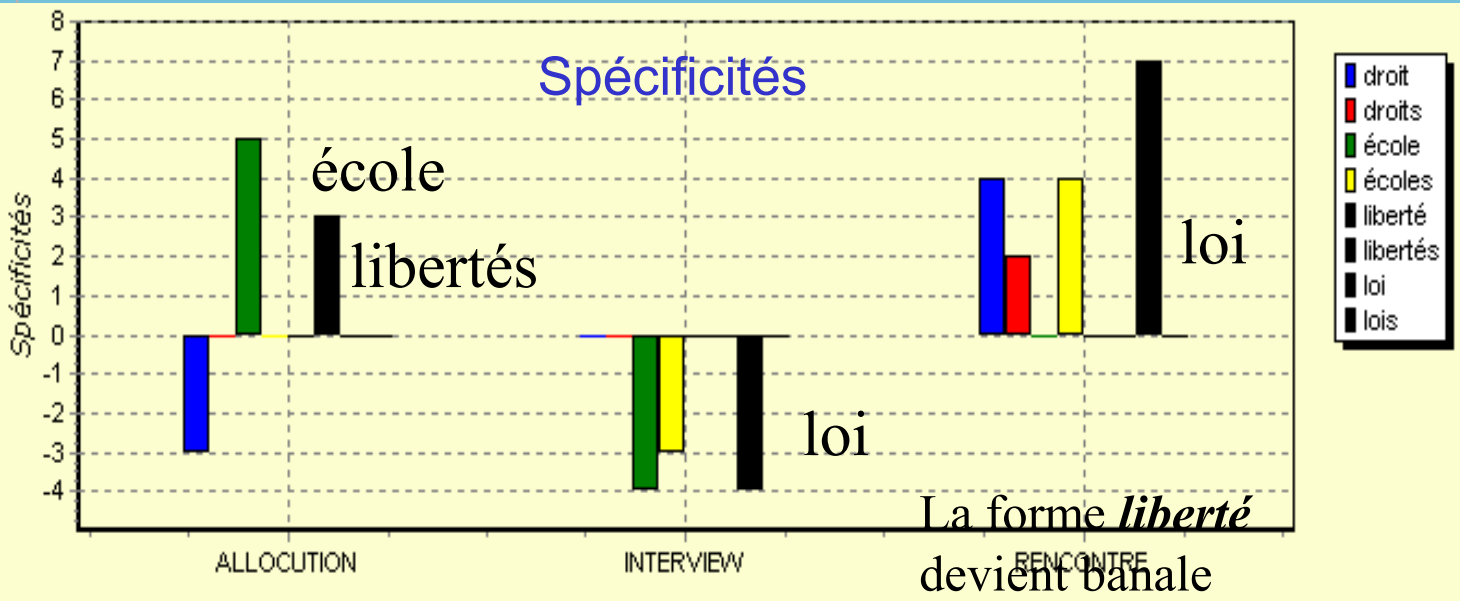
CIR Coefficient d'Implication Réciproque : rapport nombre de cooccurrences de 2 ou plusieurs formes au produit de leur fréquence

LIBERTES	CIR%	LS%	LS et L	L%	LIBERTE	CIR%
USAGE	4 2,8571	4,7619	PRIVILEGES	3 2,3810	ENSEIGNEMENT	5 0,1353
SOLIDARITE	1 0,5102	0,8547	DROITS	2 0,5556	LOIS	3 0,0999
ORDRE	1 0,2976	3,0303	HOMME	4 1,5504	ETAT	3 0,0293
NATIONALE	1 0,1832	6,0606	PENSE	4 2,0833	CONQUETES	3 0,6494
SAUVEGARDE	1 1,7857	2,0833	POUVOIR	2 0,9009	BESOIN	3 0,0248
EPANOUISSEMENT	1 1,4286	6,6667	ENSEIGNER	2 1,2821	<u>LIBERTES</u>	2 0,1855
BAFOUES	1 3,5714	3,2258	OBJECTIF	2 1,0638	CREATRICES	2 1,2987

*ne classe pas selon le même ordre les cooccurrences communes à liberté et libertés.
Enseigner, privilèges, objectif pour libertés; Privilèges, pense, homme pour liberté*

11.00 1 1 INTERVIEW LIBERATION Paris Jeudi 10 mai 1984 "Le scepticisme ambiant sur les chances de réussite de toute politique économique, je dis bien de toute politique, dure depuis bientôt dix ans. Il a atteint, puis éliminé la majorité précédente. Il frappe l'actuelle majorité. A cela une explication : l'opinion s'est lassée d'attendre le "bout du tunnel", comme cela lui fut naguère imprudemment promis, illusion dont nous ne nous sommes pas nous-mêmes suffisamment dépris. Elle veut juger sur pièces. Et je pense qu'elle a raison. Vous me demandez si je suis optimiste ou pessimiste. L'optimisme honnête se fonde sur le fait que la plupart des paramètres classiques qui permettent aux spécialistes d'apprécier l'évolution d'une situation économique sont meilleurs aujourd'hui qu'ils n'étaient pendant la première partie de ma présidence, meilleure que sous le septennat précédent. Voyez ce qui se passe pour l'inflation, le commerce extérieur, l'investissement industriel. Nous avons hérité d'une inflation à 14 % en 1981. En trois exercices nous l'avons ramenée, selon les prévisions de l'INSEE, à un rythme de 6,5 % avec objectif de 5 % ce qui constitue, si ce n'est le niveau le plus bas depuis 1971. Le commerce extérieur accusait, en 1980, un déficit de 61 milliards de francs, le verse qu'il n'atteint pas 25 milliards cette année et que nous réalisons toujours selon l'INSEE, l'équilibre l'an prochain. Cela grâce, notamment au remarquable effort de nos exportateurs. Quant à l'investissement industriel, il est enfin reparti : on prévoit 11 % d'augmentation pour 1984. Du jamais vu depuis dix ans. Bref, la France bouge et recommence à regarder droit devant elle. Le pessimisme honnête s'appuie sur l'accroissement de la dette extérieure, la lourdeur de nos importations, le taux élevé (mais il baisse) de l'argent, la faiblesse de la croissance (mais en 1981 et 1982 nous avons la plus forte d'Europe). Je considère ces problèmes avec le sérieux qu'ils méritent et je ne mésestime pas les obstacles à vaincre. Mais le règlement de la dette est à notre portée. Le seul fait d'équilibrer notre commerce extérieur modifiera du tout au tout la tendance. Pour le reste, nous agissons énergiquement afin de corriger les défauts structurels de notre économie. Suis-je optimiste ou pessimiste ? Je me contenterai de répondre : je fais ce que je dois. Certes la crise a duré plus longtemps que ne le prévoyaient la plupart des experts en 1981. Et le vieillissement de notre appareil industriel dans d'importants secteurs tel que ceux de l'industrie lourde ou du textile était plus grave que nous ne le supposions avant de gouverner. Aussi devons-nous mener la bataille sur plusieurs fronts. La lutte contre le chômage est l'un de ces fronts. En faisant de la formation professionnelle - 800.000 travailleurs irons en stage de formation cette année - la base de notre action, c'est-à-dire en préparant hommes et femmes aux métiers d'avenir, là où l'emploi se crée, nous luttons contre le chômage. En modernisant l'industrie pour gagner de nouveaux marchés, nous luttons contre le chômage. Cela suppose, dans un premier temps, de douloureuses remises en ordre - je pense à la sidérurgie - . Mais avons-nous le droit de laisser croire que les emplois seront sauvés, là où ils sont de toute façon perdus si l'on se contente de subventionner des entreprises qui ne sont plus compétitives ? La flexibilité sociale et la modernisation technologique sont les deux clés de la sortie de crise. La rigueur n'est pas une fin en soi, mais seulement un moyen de passer la tempête. Je cherche à convaincre les Français qu'il faut serrer les dents quand on veut gagner un combat difficile. J'ajoute que la rigueur n'a de sens, c'est-à-dire ne peut mobiliser les volontés de la nation qu'à la condition d'être équitablement répartie entre les diverses couches de la société. Affaire de courage, oui, mais affaire de justice aussi. Le pays reconnaîtra, au bout du compte, que la gauche au pouvoir apporte non seulement plus d'équité sociale - ce qui n'est contesté par personne - mais aussi plus d'efficacité économique que la droite - ce qui changera en profondeur et non pas, comme trop souvent de façon épidermique, le jugement que portent sur elle les Français. Alors, soyez en sûr, les conditions d'une véritable alternance démocratique seront créées. Au lieu de conquérir le pouvoir une ou deux fois par demi-siècle, portée par de brefs mouvements d'humeur, la gauche apparaîtra comme la garantie permanente d'un bon gouvernement du pays. Et comme elle continuera d'être beaucoup plus que cela, par son projet et ses valeurs, son rayonnement durera. Je veux moderniser la France dans la justice sociale. Produire, vendre cette production, accroître nos ressources pour mieux en répartir le légitime profit, cela exige une volonté sans défaillance. On ne réforme pas les structures d'un pays, on ne réduit pas les injustices et les privilèges ancrés pendant près de deux siècles sans un effort de

Spécificités Histogramme



Spécificités indicateurs

	ALLOCUTION	S	INTERVIEW	S	RENCONTRE	S
libertés	10		1		3	
laïcité	2		8		5	
France	79		64		36	
Europe	17	-5	40	+3	22	
liberté	33		23		18	
problème	19	-3	25		24	+3
droit	18	-3	21		27	+4
politique	19	-2	31	+3	15	
enseignement	32	+3	7	-5	20	
République	36	+4	9	-4	13	

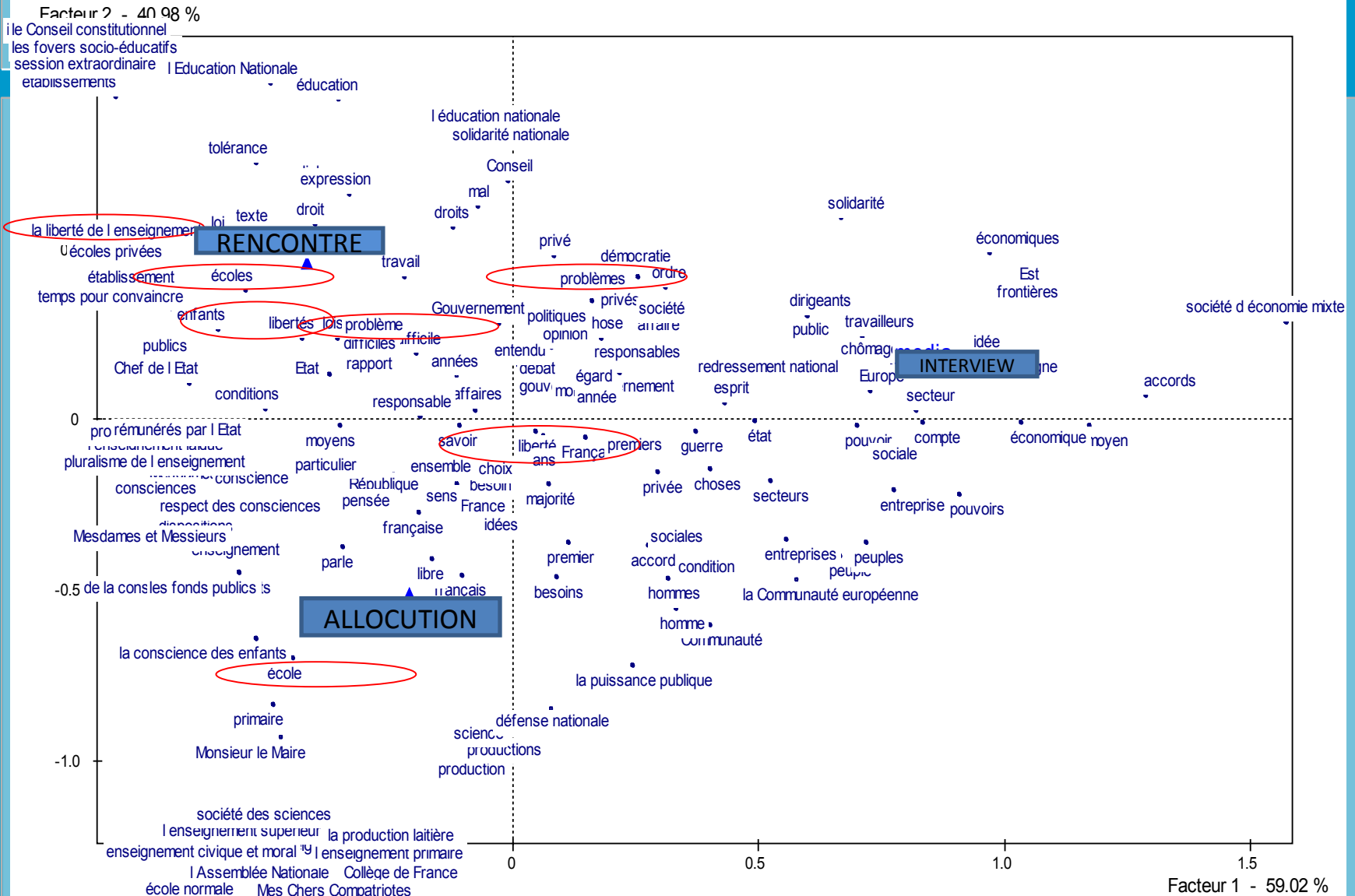
Tableau lexical de contingence

DIS : modalités

	ALLOCUTION	INTERVIEW	RENCONTRE
libertés	10	1	3
laïcité	2	8	5
France	79	64	36
Europe	17	40	22
liberté	33	23	18
problème	19	25	24
droit	18	21	27
politique	19	31	15
enseignement	32	7	20
République	36	9	13

Formes:
modalité
du corpus

Carte factorielle binaire



Analyse de Données Textuelles

- Ce type de méthodes présente l'avantage d'offrir une vision de l'élaboration de l'information par association de variables intangibles au discours.
- Un exemple concret:

OBJECTIFS:

- Cerner l'évolution des pratiques managériales dans le secteur culturel des collectivités territoriales
- Mettre en exergue les attentes des acteurs

METHODOLOGIE:

■ Exploratoire:

– trente entretiens semi-directifs dans différentes, population élus, salariés

■ Exploitation qualitative: synthèse discours avec volonté de restituer donné,

■ et analyse de données textuelles

avec le logiciel Spad

---1

« C'est comme un marionnettiste, je dis d'abord qui je suis. Augustin Cornu, je suis élu de la ville depuis 18 ans et je suis élu à la culture depuis 12 ans. C'est à dire que ça fait trois mandats que je fais à la ville et deux mandats comme élu à la culture. L'élu à la culture que je suis travaille dans le secteur culturel depuis la nuit des temps, donc je suis plus un tout jeune homme. Hein, ça fait longtemps, longtemps que je bosse dans ce secteur et il y a 12 ans, Sueur m'a dit : J'aimerais que ce soit toi qui sois élu à la culture. Je lui ai dit pourquoi pas mais tu prends un risque parce que je suis vraisemblablement un des seuls élus à la culture d'une grande ville de France qui n'a même pas son baccalauréat d'étude. Je n'ai même pas fait l'étude de la vie et j'ai appris au fil du temps parce que j'ai eu la chance d'ouvrir mes oreilles et de rencontrer des gens assez passionnants qui m'ont aidé à grandir. Donc j'ai grandi et maintenant j'essaie de redonner ce que je sais. Parce que j'ai aussi quelques responsabilités locales étant Président des adjoints à la culture des grandes villes de France. Donc, c'est une fonction que j'ai depuis trois ans, qui est assez passionnante parce que ça me permet de rencontrer et de confronter des expériences différentes suivant les localités dans lesquelles ils se trouvent. Les tailles de ville, les sociologies... enfin, tout ceci est intéressant. » Alors, mon rôle dans une collectivité, d'abord, donc je suis l'élu de la ville, il se trouve que je suis l'élu à la culture mais je suis avant tout un élu politique. Si je suis élu d'une ville comme Orléans, c'est parce que politiquement j'ai choisi le camp du mouvement et non pas le camp un petit plus calmos. Donc, tout ça fait que, la manière dont on vit dans cette ville et les relations que j'ai pu nouer au fil du temps, au fil des ans, me permettent de connaître plein de gens différents, d'origine différente, de culture différente. Et tout ça, comment faire rentrer ça dans une ville ? Comment faire en sorte que tout le monde soit reconnu ? Comment faire en sorte

RESULTATS:

- Préoccupations dues à la
 - formation, nouvelles possibilités
 - complexité administrative accrue
 - gestion statutaire, obligation de réussite au concours
- Attachement à leurs fonctions et intérêt pour leur mission de service public
- Identification de deux groupes en fonction du rapport avec le public

L 'analyse textuelle:

■ Importation et formatage des entretiens afin d 'obtenir une base exploitable par le logiciel utilisé, les entretiens sont alors assimilés à des questions ouvertes

■ Les variables:

- quantitative: niveau de formation
- qualitatives:
 - statut avec les modalités agent, directeur opérationnel, directeur administratif, élu
 - villes avec les modalités VI01 à VI06
- textuelles: entretiens

L 'analyse textuelle:

- Une analyse descriptive: calcul de fréquence des mots et segments répétés
- Edition d 'un tableau lexical de contingence
- Recherche de vocabulaire spécifique pour chaque classe (celles-ci étant définies par le statut)

L 'analyse textuelle:

tableau lexical de contingence

- Occurrences croisées entre les modalités de la variable statut et le vocabulaire

- Exemples d 'associations caractéristiques des fonctions:

- élu-théâtre, agent-légal, DO-école, DA-culturels

Mot	élu	DA	Agent	DO ¹
Accueil	3	2	9	11
Adjoint	10	2	4	32
Administratif	0	7	11	16
Animation	13	0	5	34
Association	9	1	5	13
Culturels	3	15	4	12
École	20	3	3	138
Elus	2	5	1	37
Gérer	5	2	6	30
Hein	5	0	25	11
Légal	0	0	13	0
Métier	0	1	15	23
Projet	18	11	2	39
Politique	17	8	8	30
Spectacles	22	1	4	3
Théâtre	35	0	0	17
Travailler	6	5	18	55

- D 'où volonté politique de développer l 'accès à la culture, poids de la réglementation pour le personnel non cadre, préoccupation de la diffusion du savoir pour les cadres culturels, ouverture sur les problèmes de la culture pour les responsables administratifs.

¹ L'effectif des directeurs opérationnels est plus élevé que les autres, ceci explique les fréquences élevées.

L 'analyse textuelle:

■ Deux analyses des correspondances:

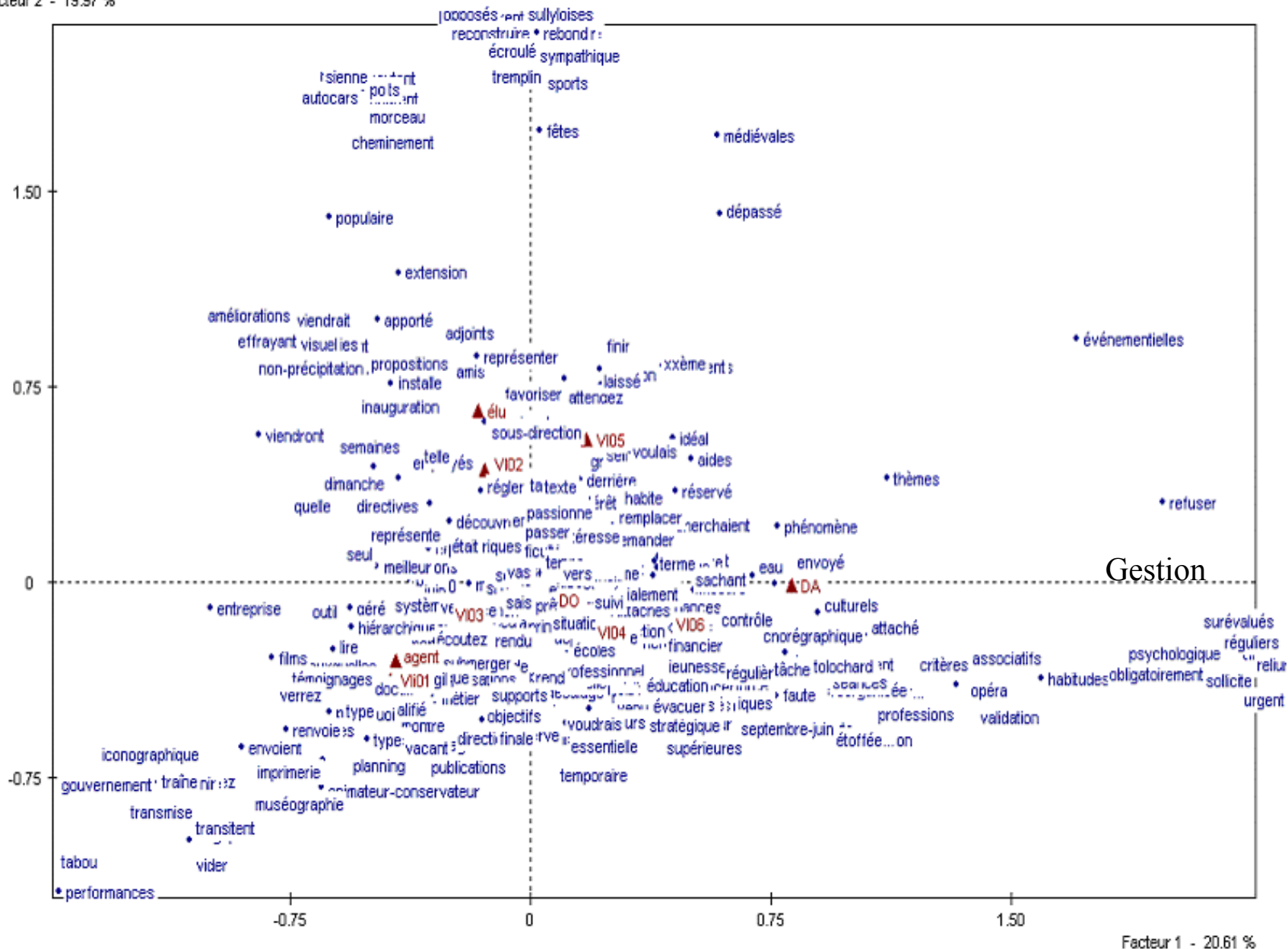
- l 'une, pour laquelle le statut et l 'ensemble du vocabulaire constituent les variables
- l 'autre, où le vocabulaire a été réduit, les mots contribuant de façon la plus significative aux axes 1 à 3 et spécifiques à la GRH dans la fonction publique sont seuls retenus .

L 'analyse textuelle:

- L 'aspect quantitatif montre que le vocabulaire reste très territorial, parmi les fréquences les plus élevées se trouvent gens (264 occurrences), service (105), public (85), travailler (84), concours (69)...
- Les néologismes semblent inexistants, les termes anglo-américains sont exceptionnels! Volonté délibérée ou type d 'organisation spécifique?

Facteur 2 - 19.97 %

•activités



•Tout le vocabulaire

Le vocabulaire dans sa globalité

- sur le premier axe la qualité de la représentation est supérieure pour le vocabulaire lié à la gestion des collectivités comme: déléguer, contrôle, charges, budgétaire, géré, habitudes, services, techniques, validation, organisation, élections, etc.
- Le deuxième est plutôt caractérisé par les activités avec des mots tels que: gratuit, sport, sponsor, spectacles etc.

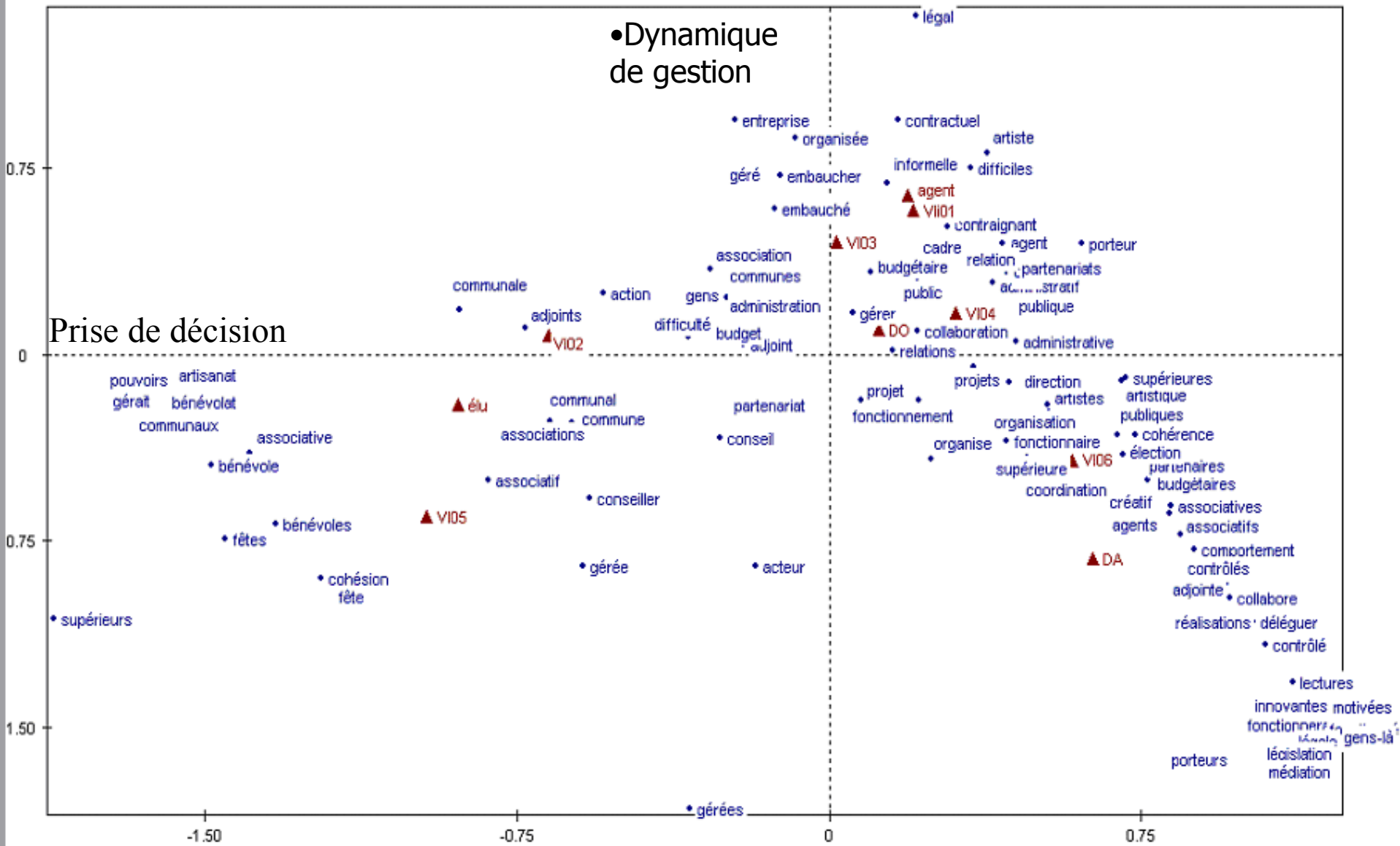
Le vocabulaire dans sa globalité

- Le troisième matérialise les relations entre le secteur culturel et ses partenaires externes sous-préfecture.., et l'environnement (économie, écologie..)
- Sur le graphique, la proximité d'un mot (modalité des variables textuelles) avec une modalité de statut traduit un usage plus fréquent dans cette sous-population de ce mot (angle aigu même quart plan)

Le vocabulaire dans sa globalité

- Les axes factoriels séparent la population en deux catégories, l'une composée des acteurs en relation directe avec le public: les élus, les agents, l'autre regroupe les directeurs opérationnels, les directeurs administratifs.
- En ce qui concerne la séparation horizontale, les administratifs se trouvent face aux élus

Facteur 2 - 23.88 %



•Réduction du vocabulaire

Vocabulaire restreint à la GRH

- Premier axe caractérisé par la prise de décision et sa mise en œuvre:
 - action, direction, projets, association, fêtes, bénévolat...
- Deuxième axe reflète la dynamique de la gestion:
 - gens, porteur, innovante, médiation, relation, associatif...

Vocabulaire restreint à la GRH

- Une séparation entre le sous-système opérationnel (agent-DO) et le sous-système de décision (élu-DA) apparaît.
- La modalité *élu* reste différenciée des salariés: Les politiques face aux administratifs.

Vocabulaire restreint à la GRH

- La modalité élu est très proche des mots comme gérant, pouvoirs, bénévolat, associative, fêtes, conseil, cohésion... qui traduisent la prise de décision et les relations extérieures à l'organisation territoriale.
- Le vocabulaire des DO et DA soulignent l'importance des contraintes budgétaires et légales; l'importance des relations avec le public pour les DO, les agents

Conclusion

- L'attachement des différents acteurs à leurs missions est à souligner, tant l'aspect technique que relationnel.
- Le statut et la mission induisent une relative spécificité des termes et des préoccupations.
- Une analyse lexicale approfondie offrirait la possibilité de tester la stabilité de cette typologie afin de mieux appréhender les attentes des acteurs face aux changements.

Exploiter les Données Textuelles

- Ce type de méthodes présente l'avantage d'offrir une vision de l'élaboration de l'information par association de variables intangibles au discours.*
- La complexité: élaboration de variables, unités statistiques, matrice creuse, apport externe, pose le problème de stabilité de la structure engendrée, voire de validité
- Comparer le corpus brut au corpus lemmatisé ou réduit en fonction des objectifs
- Rapprocher le corpus "entretiens" des caractéristiques des individus

* cf LEBART L.

Bibliographie...

GAUSSIÉ Y., YVON F., Modèles statistiques pour l'accès à l'information textuelle, Lavoisier, 2011

LEBART L., SALEM A., Statistique textuelle, Dunod, 1994

LEBART L., MORIN A., Base de données et statistiques, Dunod, 2002

REINERT M., “ Approche statistique et problème du sens dans une enquête ouverte”, *Journal de la Société Française de Statistique*, tome 142, vol 4, 2001

www.cavi.univ-paris3.fr/lexicometrica

Actes des JADT 2002, 2004, 2006, 2008, 2010, 2012 ...

Logiciels téléchargeables gratuits: LEXICO2, TXM

LEBART L., l'articulation entre exploration et inférence en analyse statistique de textes, *JADT 2012*