

Lexicometrics

a.k.a

Textual Statistics

CONTRAS

- A counter-intuitive approach
- Where's the 'reference' dictionary ?
- A text is not an equi-probable population!

PROS

- Exhaustive, objective and fast

Statistics & Probabilistic Stats ?

In general

Descriptive Statistics - How many are there ?

Inferential Statistics - How many should there be ?

Applied to text corpora

- **Observed** frequencies versus **theoretical** frequencies
- Endogeneous Norm

Text sampling

CORPUS

ECHANTILLON

mes chers concitoyens,
depuis vingt-cinq ans, on vous dit que contre le chômage,
la crise du logement, l'exclusion, les délocalisations, on
ne peut rien faire. on vous explique que la politique ne
peut plus agir, qu'avec la mondialisation les états n'ont
plus de rôle à jouer. je vous propose d'en finir avec ce
renoncement, avec cette résignation.

je vous propose d'en finir avec ce
renoncement, avec cette résignation.

depuis 2002, je vous ai montré que l'on pouvait agir. bien
sûr, il reste encore beaucoup à faire. mais, en cinq ans, la
délinquance a reculé et l'immigration a été mieux
maîtrisée.

depuis 2002, je vous ai montré que l'on pouvait agir.

je sais que si nous voulons, nous pouvons. notre pays
traverse une triple crise. une crise du travail: il n'est plus
assez récompensé. une crise de l'action politique: elle
n'est plus efficace. une crise d'identité : notre nation n'ose
plus défendre, ni transmettre ses valeurs.

je sais que si nous voulons, nous pouvons.

il n'y a aucune fatalité. notre pays peut changer. il peut
relever les défis qui sont les siens. il peut résoudre les
problèmes qui sont les vôtres, tout en restant lui-même.

ces dernières semaines, j'ai eu l'occasion de vous
présenter mes propositions. je n'ai pas cherché à vous
cacher la vérité, parce que je pense que la vie politique de
notre pays a trop longtemps souffert d'un manque de
sincérité. mais, au-delà des mesures concrètes que je
prendrai si je suis élu, à l'heure où vous devez choisir
celui qui va conduire notre pays pendant cinq ans, je veux
surtout que vous sachiez quelles sont les valeurs qui
guideront mon action.

je n'ai pas cherché à
vous cacher la vérité, parce que je pense que la vie
politique de notre pays a trop longtemps souffert d'un
manque de sincérité. mais, au-delà des mesures concrètes
que je prendrai si je suis élu, à l'heure où vous devez
choisir celui qui va conduire notre pays pendant cinq ans,
je veux surtout que vous sachiez quelles sont les valeurs
qui guideront mon action.

je veux d'abord être le président de la valeur travail

(...)

je veux d'abord être le président de la valeur travail

(...)

Theoretical Frequency

- T : total number of words in corpus
- t : total number of words in sample
- F : total frequency of word X in corpus
- f : total frequency of word X in sample

Compared to observed frequency:

More, less or equal

Evolving preoccupations

- 1970 Literary studies
- 1980 Political Discourse
- 1990 Data mining, mainly processing polls
- 2000 Web mining

Lexicometric methods

- Multidimensional Statistics
 - Factor Analysis
 - Hierarchical Clustering
- Specificities (Contrastive Statistics)
- Syntagmatic Statistics

Co-occurrence approaches

- Specific co-occurrences
- Poly-co-occurrences
- Co-occurrence threads
- Hyper-threads

A corpus...

Political programs of the candidates for the French Presidential elections of may 2007



Philippe de Villiers
Ancien Ministre,
Président du Conseil Général de la Vendée

La fierté d'être Français



FRANÇOIS BAYROU
LA FRANCE DE TOUTES NOS FORCES

Je serai le président qui réunira les Français pour la France, pour la France qui vivra ensemble, Français de tous les jours.



VOTEZ LE PEN

Vive la Vie, Vive la République, Vive la France, Vive le Pays.

LE PEN
Président de la République



La France Présidente


Le changement

Ségolène Royal
Election Présidentielle du 22 avril 2007



Olivier BESANGENOT
Election Présidentielle du 22 avril 2007

Nos vies valent plus que leurs profits!



ENSEMBLE TOUT DEVIENT POSSIBLE


NICOLAS SARKOZY
sarkozy.fr



ENSEMBLE TOUT DEVIENT POSSIBLE

NICOLAS SARKOZY

Je suis Nicolas Sarkozy, je suis Français, je suis président de la République.



ET TOUT DEVIENT POSSIBLE

Je veux être le Président du pouvoir d'achat

Je veux être le Président qui garantit de nouvelles protections

Je veux être le Président d'un Etat fort et impartial



ENSEMBLE TOUT DEVIENT POSSIBLE

Je veux d'abord être le Président de la valeur travail

NICOLAS SARKOZY
sarkozy.fr

12 candidates
4 pages per candidate
20 000 occurrences

Co-occurrences in the dictionary ?

Rang	Forme	Fréq.	Rang	Forme	Fréq.	Rang	Forme	Fréq.
1	<i>de</i>	1 024	31	<i>c</i>	66	61	<i>publics</i>	35
2	<i>la</i>	653	32	<i>notre</i>	66	62	<i>europa</i>	35
3	<i>les</i>	609	33	<i>a</i>	64	63	<i>ils</i>	35
4	<i>et</i>	606	34	<i>tous</i>	64	64	<i>toutes</i>	35
5	<i>des</i>	501	35	<i>politique</i>	62	65	<i>comme</i>	34
6	<i>l</i>	459	36	<i>nos</i>	60	66	<i>sans</i>	34
7	<i>le</i>	395	37	<i>sont</i>	60	67	<i>droits</i>	33
8	<i>à</i>	379	38	<i>s</i>	56	68	<i>européenne</i>	33
9	<i>d</i>	281	39	<i>n</i>	55	69	<i>elle</i>	32
10	<i>pour</i>	278	40	<i>france</i>	54	70	<i>emploi</i>	32
11	<i>en</i>	263	41	<i>ans</i>	51	71	<i>faut</i>	32
12	<i>du</i>	213	42	<i>contre</i>	51	72	<i>son</i>	32
13	<i>est</i>	186	43	<i>qu</i>	51	73	<i>vote</i>	32
14	<i>une</i>	186	44	<i>ou</i>	49	74	<i>leur</i>	31
15	<i>qui</i>	176	45	<i>pays</i>	49	75	<i>doit</i>	30
16	<i>je</i>	162	46	<i>ce</i>	48	76	<i>services</i>	30
17	<i>un</i>	162	47	<i>tout</i>	48	77	<i>pouvoir</i>	29
18	<i>dans</i>	149	48	<i>vie</i>	48	78	<i>cette</i>	28
19	<i>que</i>	142	49	<i>ont</i>	47	79	<i>emplois</i>	28
20	<i>au</i>	113	50	<i>se</i>	46	80	<i>entre</i>	28
21	<i>plus</i>	112	51	<i>travail</i>	44	81	<i>j</i>	28
22	<i>il</i>	101	52	<i>faire</i>	42	82	<i>veux</i>	28
23	<i>nous</i>	101	53	<i>mais</i>	41	83	<i>monde</i>	27
24	<i>aux</i>	99	54	<i>entreprises</i>	40	84	<i>république</i>	27
25	<i>sur</i>	96	55	<i>être</i>	40	85	<i>gauche</i>	26
26	<i>avec</i>	95	56	<i>ses</i>	40	86	<i>on</i>	26
27	<i>par</i>	88	57	<i>sociale</i>	40	87	<i>salaires</i>	26
28	<i>ne</i>	77	58	<i>français</i>	38	88	<i>depuis</i>	25
29	<i>vous</i>	75	59	<i>droit</i>	37	89	<i>loi</i>	25
30	<i>pas</i>	71	60	<i>ces</i>	36	90	<i>sa</i>	25

ump

fn

deVILLIERS

udf

ps

pc

cpnt

BOVE

verts

lo

lcr

SCHW

SPECIFICITES - UDF, DE VILLIERS, LE PEN, UMP

Positives

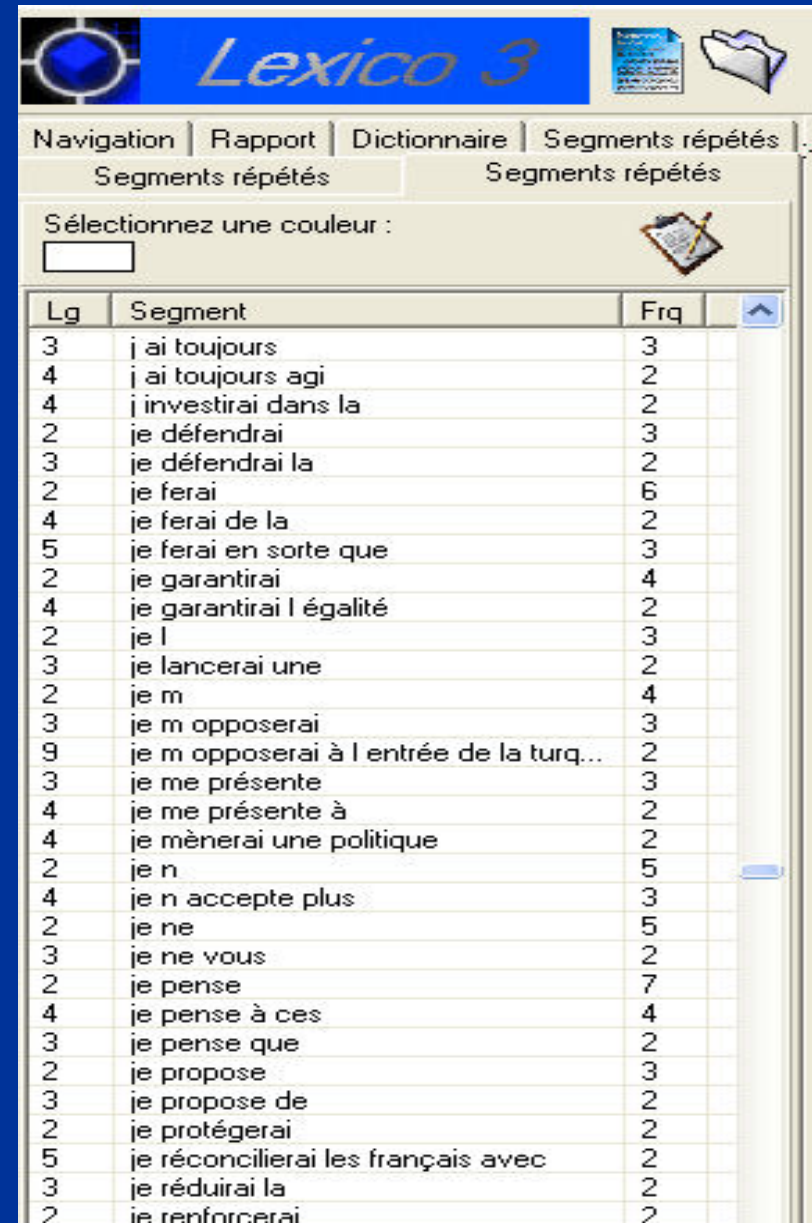
Terme	Frq Tot.	Frq P...	Spécif
je	160	120	35
français	38	32	13
VILLIERS	17	17	10
i	28	22	8
vendée	11	11	7
france	53	33	7
président	23	18	7
notre	66	36	6
veux	28	20	6
immigration	11	10	5
travail	44	23	4
le	394	141	4
plus	112	48	4
que	142	60	4
suis	10	7	3
cing	17	10	3
enfants	17	10	3
votre	22	13	3
n	55	26	3
vous	75	31	3
SARKOZY	10	7	3
pays	49	22	3
état	21	10	2
ai	17	9	2
toujours	15	8	2
mon	14	8	2
valeurs	13	8	2
peut	15	8	2
aujourd	16	9	2
hui	16	9	2
ni	15	8	2
son	32	15	2
pouvoir	29	13	2
sera	11	7	2
ans	51	22	2
françois	10	6	2
l	457	144	2
BAYROU	11	7	2
m	10	6	2

Négatives

Terme	Frq Tot.	Frq P...	Spécif
ma	16	1	-2
d	281	64	-2
et	606	143	-2
profits	19	1	-2
création	20	1	-2
avril	22	2	-2
droit	37	5	-2
elle	32	4	-2
classes	18	1	-2
pour	277	58	-3
du	213	40	-3
européenne	32	2	-3
non	22	1	-3
gauche	26	1	-3
ces	36	2	-4
s	56	5	-4
des	501	92	-7

Syntagmatic Statistics


- Repeated Segments
- Concordances
- Distributional Inventory



The screenshot shows the Lexico 3 software interface. The title bar includes the Lexico 3 logo and a folder icon. The menu bar contains 'Navigation', 'Rapport', 'Dictionnaire', and 'Segments répétés'. Below the menu bar, there are two tabs labeled 'Segments répétés'. A label 'Sélectionnez une couleur :' is followed by a color selection box and a notepad icon. The main area displays a table with three columns: 'Lg', 'Segment', and 'Frq'. The table contains 30 rows of data, each representing a different French phrase and its frequency.

Lg	Segment	Frq
3	j ai toujours	3
4	j ai toujours agi	2
4	j investirai dans la	2
2	je défendrai	3
3	je défendrai la	2
2	je ferai	6
4	je ferai de la	2
5	je ferai en sorte que	3
2	je garantirai	4
4	je garantirai l égalité	2
2	je l	3
3	je lancerai une	2
2	je m	4
3	je m opposerai	3
9	je m opposerai à l entrée de la turq...	2
3	je me présente	3
4	je me présente à	2
4	je mènerai une politique	2
2	je n	5
4	je n accepte plus	3
2	je ne	5
3	je ne vous	2
2	je pense	7
4	je pense à ces	4
3	je pense que	2
2	je propose	3
3	je propose de	2
2	je protégerai	2
5	je réconcilierai les français avec	2
3	je réduirai la	2
2	je renforcerai	2


Syntagmatic Statistics (2)

Forme: Tri: Regroupement: Largeur: 

Expression rationnelle Type de documentation: Délimiteurs de séquence:

ment . dans cette campagne , avec cpnt , je veux faire entendre la parole de toutes
on trois cent mille voix . aujourd ' hui je prends le relais avec son soutien et son
et pour les françaises et les français , je propose : - de mettre en place une vraie
ous serons déterminants au second tour . je compte sur vous , vous pouvez compter sur
es grandes choses se font avec le cœur , je me présente à vous comme " l ' homme du
nt retrouver la maîtrise de son avenir . je ne suis pas un politicien , contrairement
es . par tempérament et par expérience , je suis un créateur , un entrepreneur , un
ateur , un entrepreneur , un bâtisseur . je vous propose d ' élargir demain l ' expérience
ort de la cohérence de mes convictions , je sollicite votre confiance . " rassembler
74 ans , maire " dans mon département , je constate chaque jour la montée de l ' islamisation
préoccupations des français . en 2007 , je voterai utile , je voterai philippe de VILLIERS
français . en 2007 , je voterai utile , je voterai philippe de VILLIERS . " christophe
hristophe , 31 ans , conducteur de bus " je regrette que m . sarkozy soit favorable
votre emploi contre les délocalisations je protégerai votre travail en instaurant des
baisser les charges de 50% pour les pme je favoriserai fiscalement les petites entreprises
enter le pouvoir d ' achat des français je libérerai votre pouvoir d ' achat en supprimant
partie des députés à la proportionnelle je réconcilierai les français avec la politique
le respect et le patriotisme à l ' école je rétablirai la discipline , le respect ,
romouvoir les repères civiques et moraux je lancerai une grande politique familiale
ncerai une grande politique familiale et je tournerai la société vers les valeurs de
de la vie et de la propriété pour tous . je m ' opposerai au mariage homosexuel et à
er des emplois et de la qualité de vie . je défendrai l ' agriculture et la chasse .
défendrai l ' agriculture et la chasse . je développerai les services publics de proximité
s 35 heures : 20 milliards d ' économie je mettrai fin à l ' endettement de la france
' immigration et assimiler les étrangers je stopperai l ' immigration pour donner la
re de la présence sur notre territoire . je lancerai une politique de co - développement
onner à la france sa place dans le monde je m ' opposerai à l ' entrée de la turquie
oserai à l ' entrée de la turquie dar et je proposerai une europe fondée sur les nations
e fondée sur les nations impérialismes , je ferai de la france une puissance d ' équilibre
sonne toujours deux fois . aujourd ' hui je fais le même métier et gagne , comme la
us , moins de 1 400 euros net par mois . je me présente à nouveau à l ' élection présidentielle
papiers scolarisés dans nos écoles . . . je souhaite être le porte - voix de toutes
en 1968 , cela le reste aujourd ' hui ! je veux être le candidat qui défend l ' augmentation
' immobilier est un véritable scandale , je propose de les bloquer par la loi , de réquisitionner
élocaliser , sous peine de réquisition . je veux que tous ensemble nous luttons contre
coups de son conjoint . c ' est pourquoi je soutiens la proposition de loi - cadre contre

Syntagmatic Statistics (3)

Forme : je		Tri : Aucun		Regroupement : <Aucun>		Largeur : 40			
<input type="checkbox"/> Expression rationnelle		Type de documentation : Inventaire distributio		Délimiteurs de séquence : ...!/?_-\'"000\$\$					
	2	----	----	----	je	me	présente	à	
	2	----	----	----	je	mènerai	une	politique	
5		----	----	----	je	n			
	3	----	----	----	je	n	accepte	plus	
5		----	----	----	je	ne			
	2	----	----	----	je	ne	vous		
7		----	----	----	je	pense			
	4	----	----	----	je	pense	à	ces	
	2	----	----	----	je	pense	que		
3		----	----	----	je	propose			
	2	----	----	----	je	propose	de		
2		----	----	----	je	protégerai			
		2	----	----	je	réconcilierai	les	français	avec
	2	----	----	----	je	réduirai	la		
2		----	----	----	je	renforcerai			
6		----	----	----	je	serai			
	5	----	----	----	je	serai	le		
	4	----	----	----	je	serai	le	président	
		3	----	----	je	serai	le	président	qui
			2	----	je	serai	le	président	qui réunira les français pour les faire
3		----	----	----	je	sollicite			
	2	----	----	----	je	sollicite	vos	suffrages	
2		----	----	----	je	souhaite			
	2	----	----	----	je	soutiendrai	notre	agriculture	
9		----	----	----	je	suis			
	4	----	----	----	je	suis	le	candidat	
			3	----	je	suis	le	candidat	de la
	2	----	----	----	je	suis	un		
28		----	----	----	je	veux			
	7	----	----	----	je	veux	être	le	
		6	----	----	je	veux	être	le	président
			2	----	je	veux	être	le	président d
			2	----	je	veux	être	le	président qui

Specific Co-occurrences

Which words attract each other ? —

- A 'co-occurrence' is the apparition of two words at the same time, in the same context.

Specific Co-occurrences (2)

CORPUS

mes chers concitoyens,
depuis vingt-cinq ans, on vous dit que contre le chômage,
la crise du logement, l'exclusion, les délocalisations, on
ne peut rien faire. on vous explique que la politique ne
peut plus agir, qu'avec la mondialisation les états n'ont
plus de rôle à jouer. je vous propose d'en finir avec ce
renoncement, avec cette résignation.

depuis 2002, je vous ai montré que l'on pouvait agir. bien
sûr, il reste encore beaucoup à faire. mais, en cinq ans, la
délinquance a reculé et l'immigration a été mieux
maîtrisée.

je sais que si nous voulons, nous pouvons. notre pays
traverse une triple crise. une crise du travail: il n'est plus
assez récompensé. une crise de l'action politique: elle
n'est plus efficace. une crise d'identité : notre nation n'ose
plus défendre, ni transmettre ses valeurs.

il n'y a aucune fatalité. notre pays peut changer. il peut
relever les défis qui sont les siens. il peut résoudre les
problèmes qui sont les vôtres, tout en restant lui-même.

ces dernières semaines, j'ai eu l'occasion de vous
présenter mes propositions. je n'ai pas cherché à vous
cacher la vérité, parce que je pense que la vie politique de
notre pays a trop longtemps souffert d'un manque de
sincérité. mais, au-delà des mesures concrètes que je
prendrai si je suis élu, à l'heure où vous devez choisir
celui qui va conduire notre pays pendant cinq ans, je veux
surtout que vous sachiez quelles sont les valeurs qui
guideront mon action.

je veux d'abord être le président de la valeur travail
(...)

ECHANTILLON

je vous propose d'en finir avec ce
renoncement, avec cette résignation.

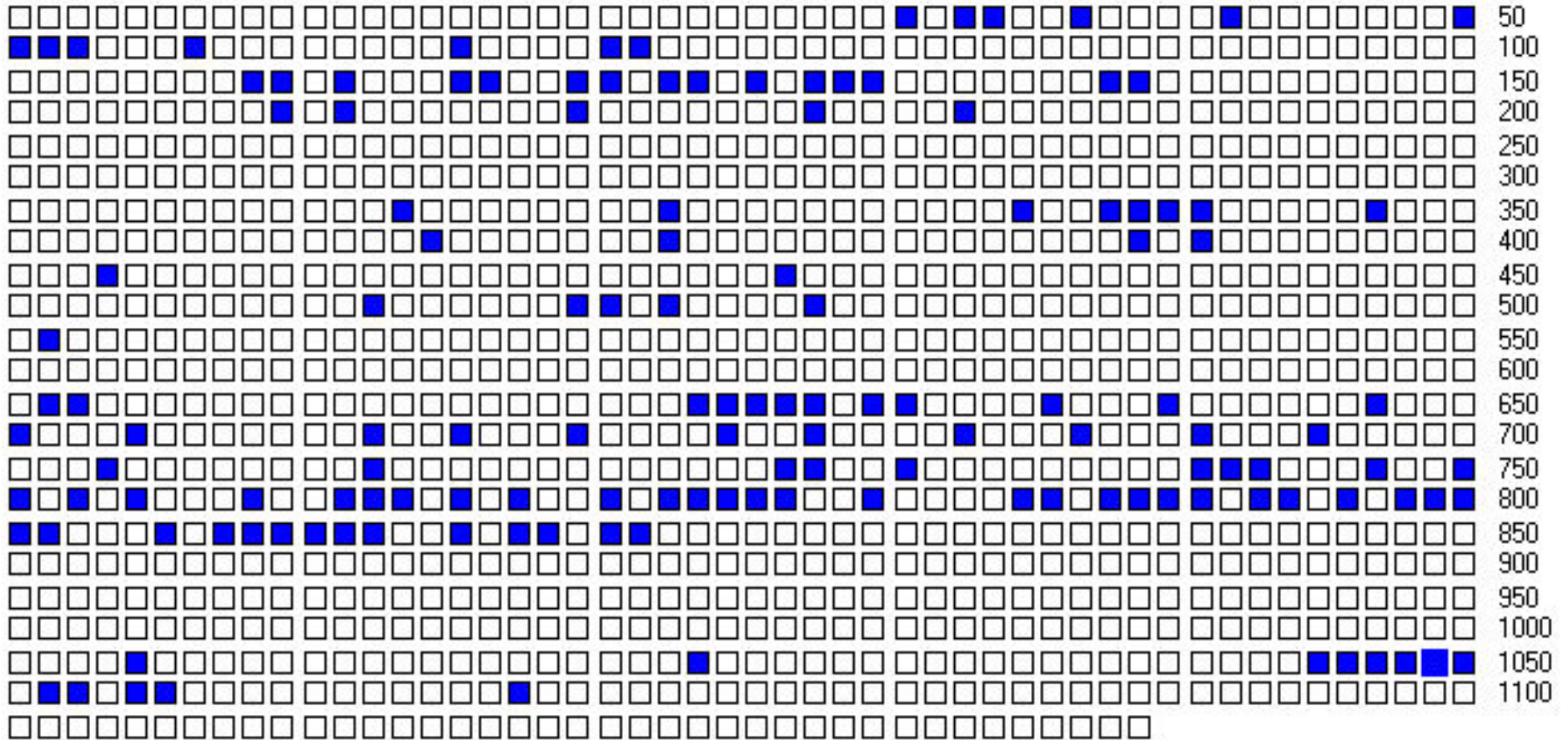
depuis 2002, je vous ai montré que l'on pouvait agir.

je sais que si nous voulons, nous pouvons.

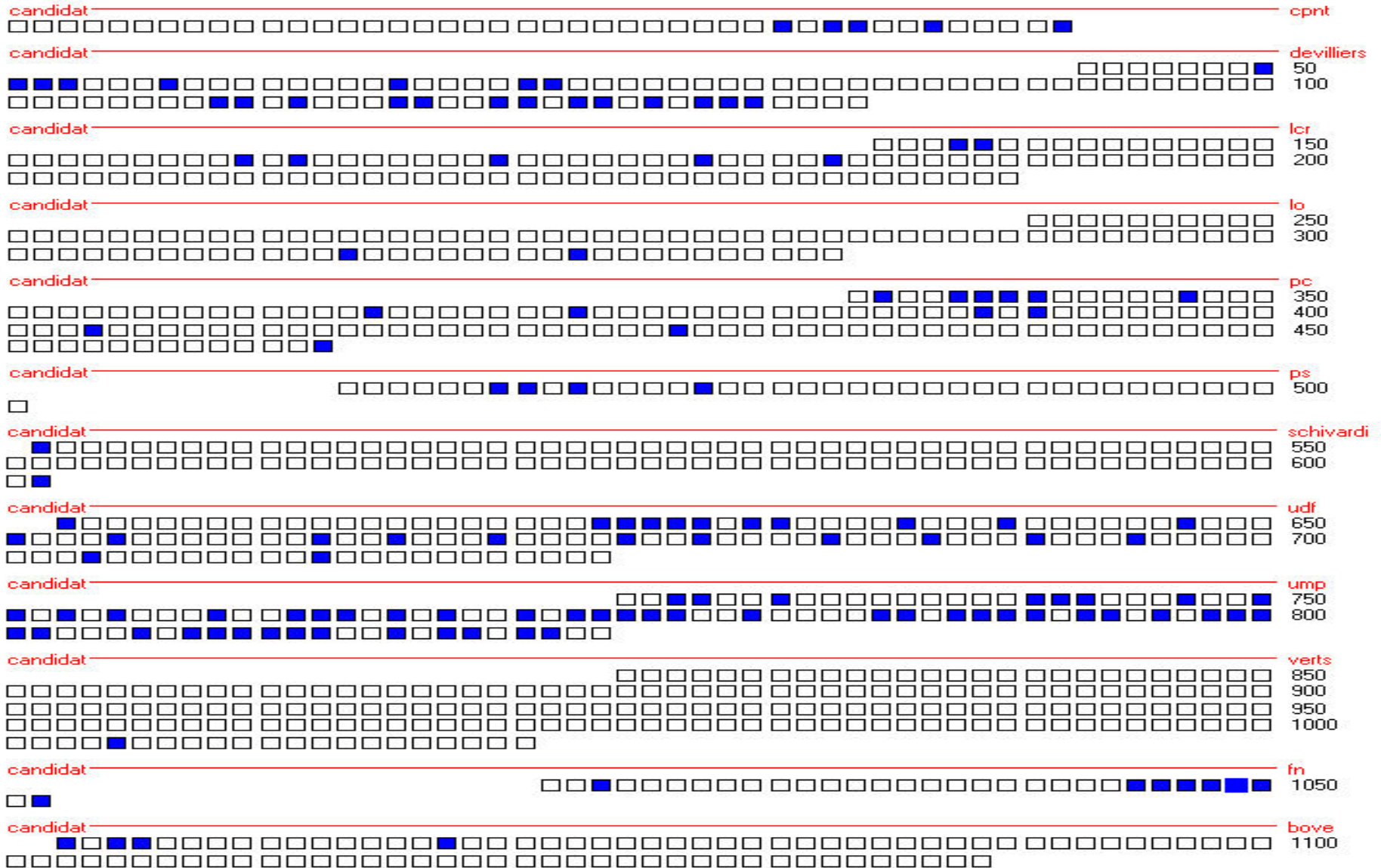
je n'ai pas cherché à
vous cacher la vérité, parce que je pense que la vie
politique de notre pays a trop longtemps souffert d'un
manque de sincérité. mais, au-delà des mesures concrètes
que je prendrai si je suis élu, à l'heure où vous devez
choisir celui qui va conduire notre pays pendant cinq ans,
je veux surtout que vous sachiez quelles sont les valeurs
qui guideront mon action.

je veux d'abord être le président de la valeur travail
(...)

Text Mapping: *je*



Text Mapping: *je* (2)



Specific Co-occurrences (4)

From keyword *je*
(142 sentences = 2 890 occ.)

N°	Forme	Fréq.	Co-fréq.	Spécificité	Contextes
1	veux	28	28	+24	26
2	que	142	49	+9	39
3	suis	10	10	+9	10
4	vous	75	30	+8	26
5	pense	8	7	+6	7
6	ferai	6	6	+6	6
7	serai	6	6	+6	6
8	notre	66	23	+5	20
9	me	7	6	+5	6
10	du	213	9	-6	9

Poly-cooccurrences








































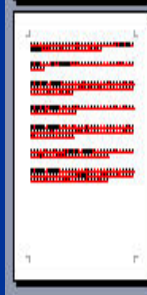











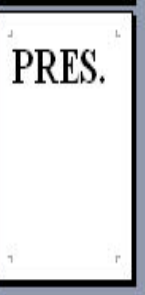
Beyond binary co-occurrence ...?

Revealing complex lexical systems

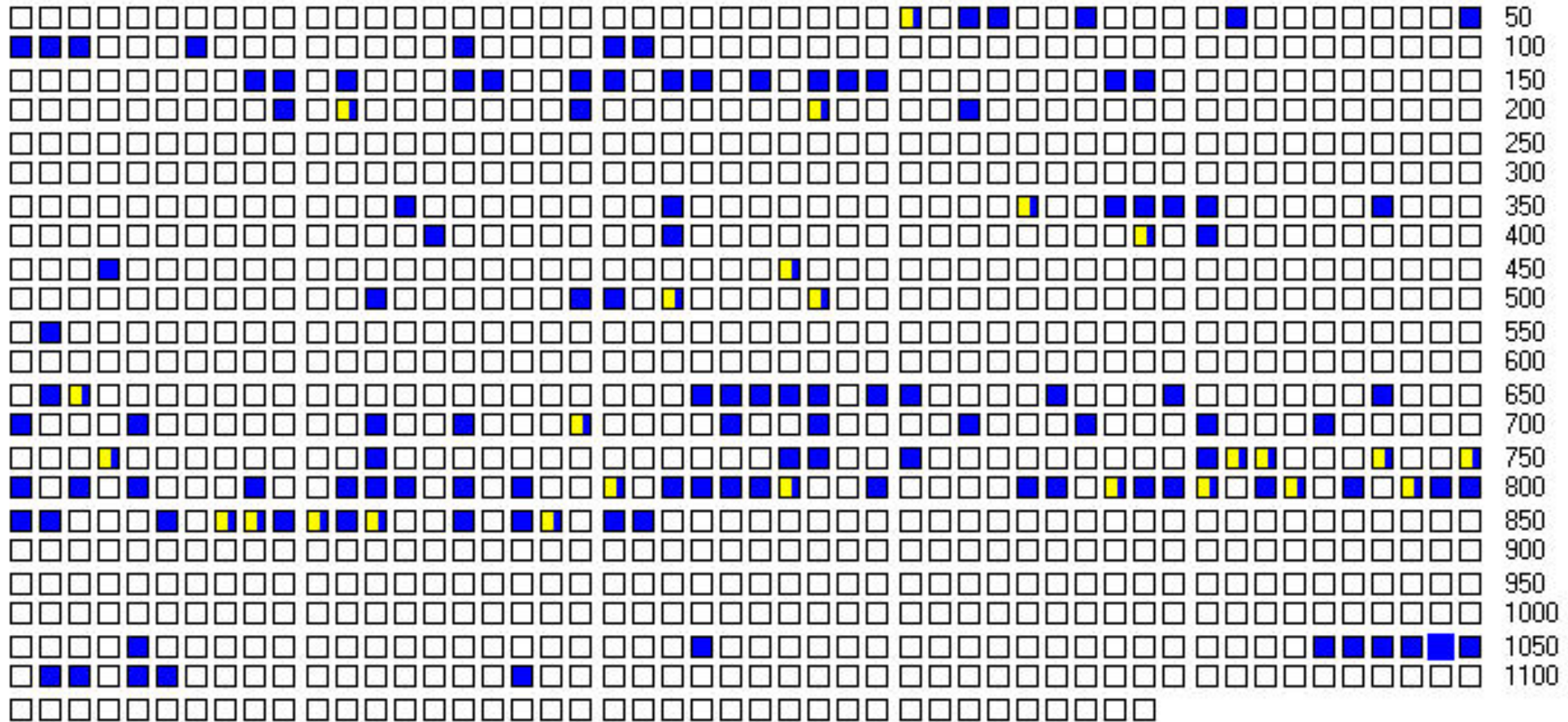
Method : Reiterate co-occurrence analysis so as to determine multiple and simultaneous word attractions.

Ex. : On the basis of pairs AB et BC, establish poly-co-occurrence ABC.

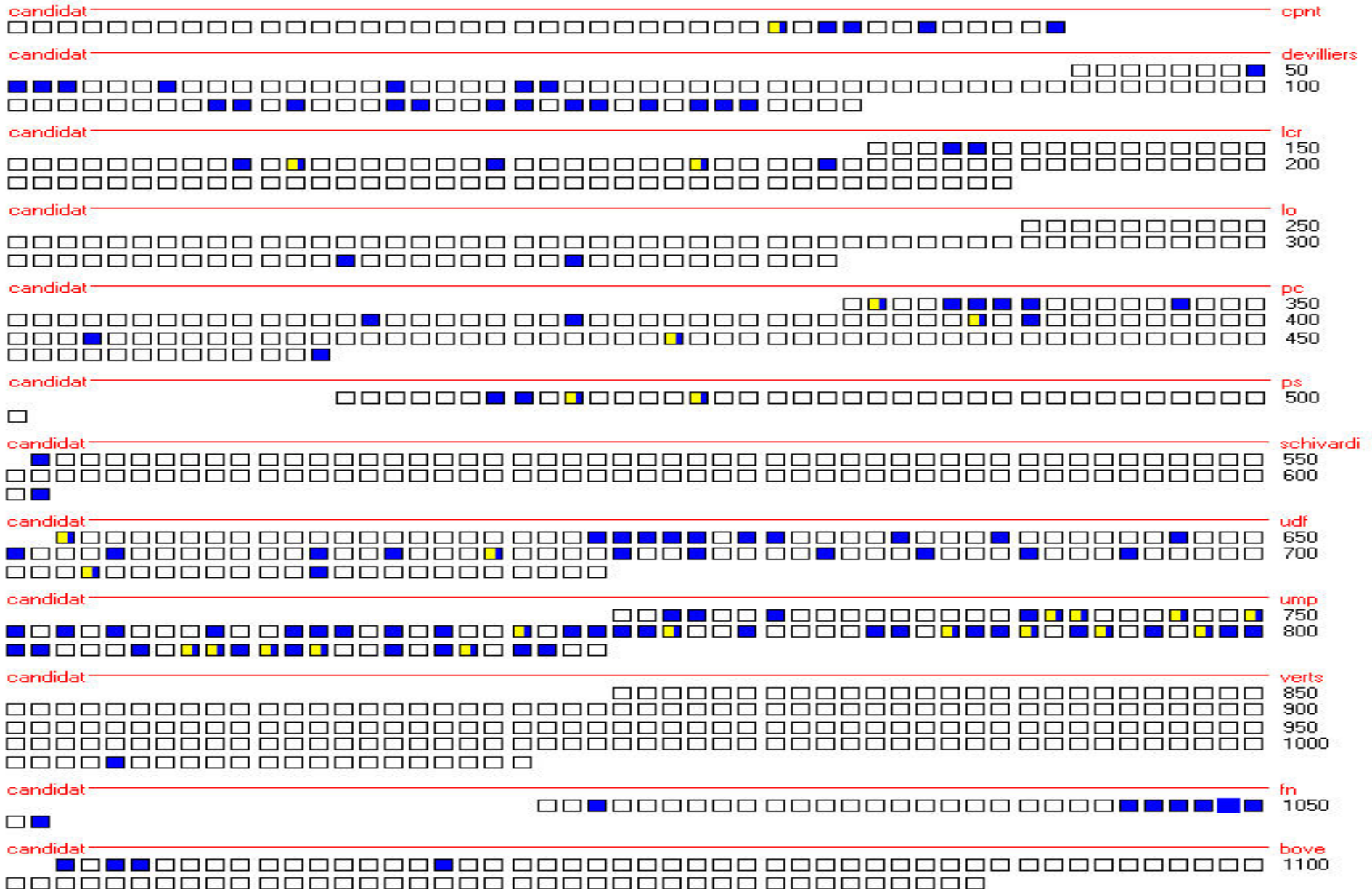
Re-sampling data

Text Mapping: *je+veux*



Text Mapping: *je+veux* (2)



Which are the co-occurrences of *je* ?

N°	Forme	Fréq.	Co-fréq.	Spécificité	Contextes
1	veux	28	28	+24	26
2	que	142	49	+9	39
3	suis	10	10	+9	10
4	vous	75	30	+8	26
5	pense	8	7	+6	7
6	ferai	6	6	+6	6
7	serai	6	6	+6	6
8	notre	66	23	+5	20
9	me	7	6	+5	6
10	du	213	9	-6	9

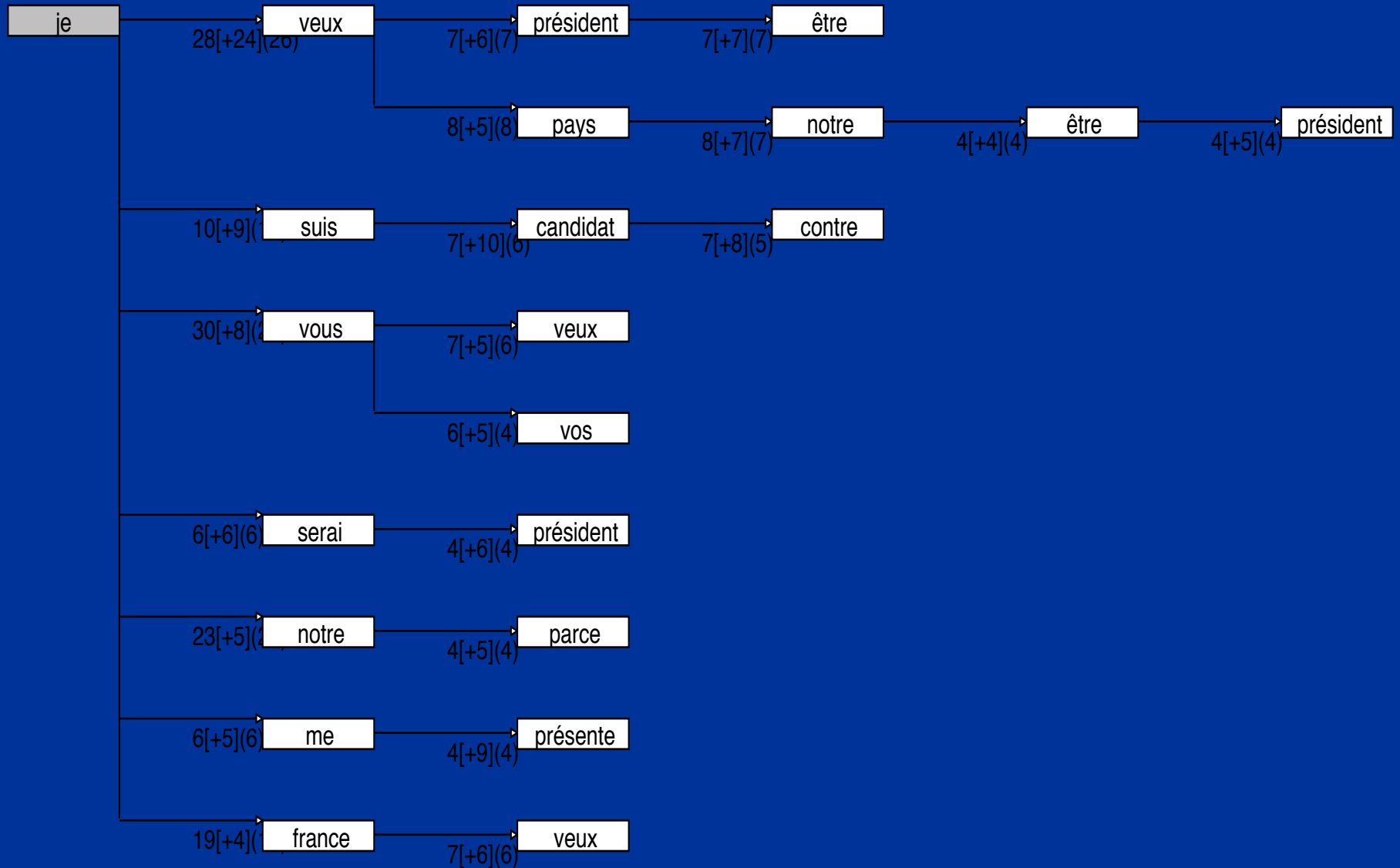
je + *veux* ?

N°	Forme	Fréq.	Co-fréq.	Spécificité	Contextes
1	que	142	16	+6	13
2	président	23	7	+6	7
3	pays	49	8	+5	8
4	être	40	8	+5	8

je + *veux* + *président* ?

N°	Forme	Fréq.	Co-fréq.	Spécificité	Contextes
1	être	40	7	+7	7

Poly-co-occurrences (3)



Back to context

UDF

je veux être le président de la république qui rendra son équilibre à notre pays.

UMP

je veux d'abord être le président de la valeur travail ...

je veux être le président du pouvoir d'achat...

je veux être le président qui garantit de nouvelles protections ...

je veux être le président d'un état fort et impartial ...

je veux être le président qui prépare l'avenir des générations futures

...

je veux être le président d'une france fière de ses valeurs et de son identité ...

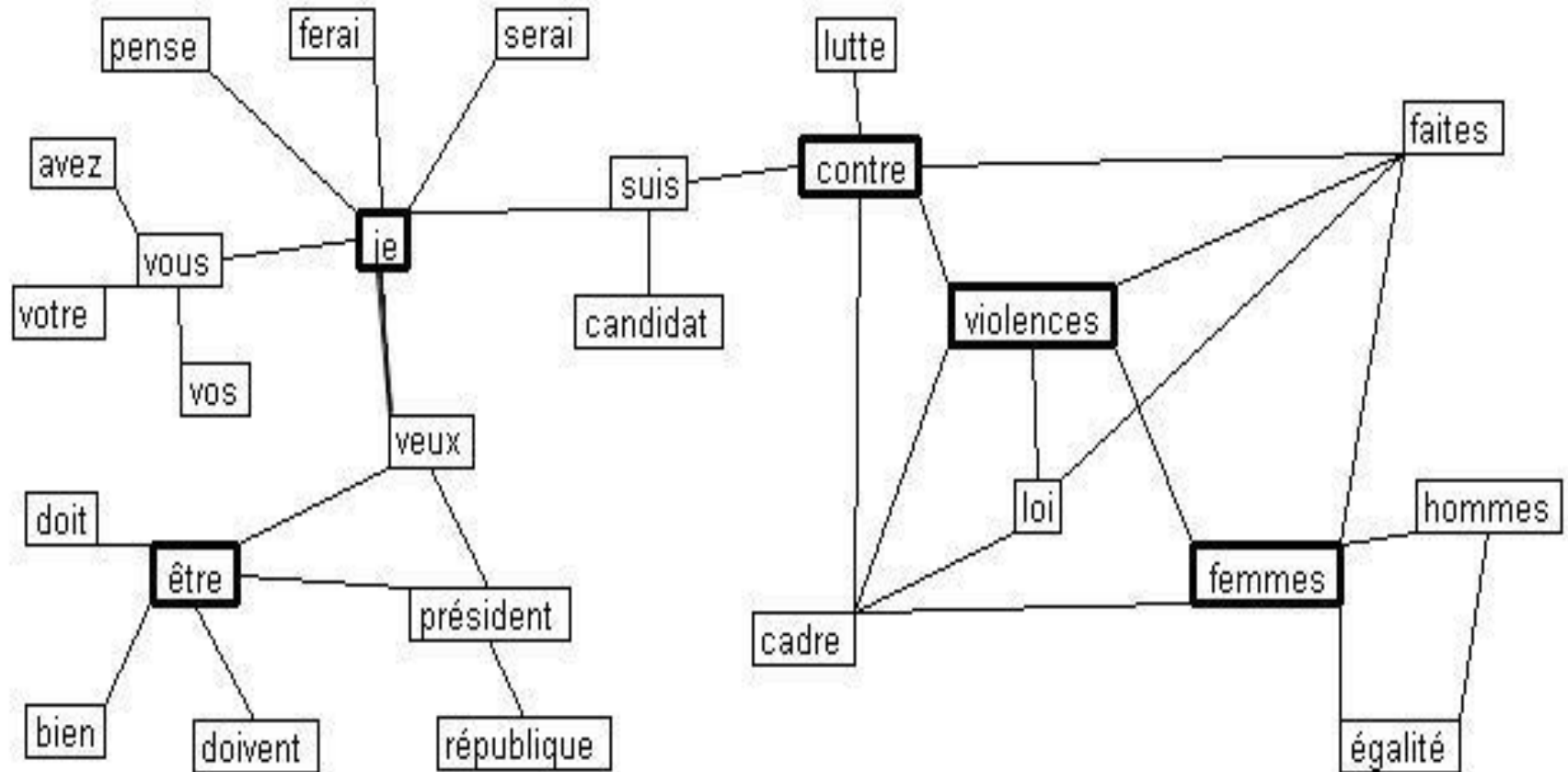
Co-occurrence Threads

- *Which network of co-occurrences structures the text?*

Method: Reiterate co-occurrence analysis to reveal binary and reciprocal co-occurrences so as to establish non-transitive associations.

Ex. : On the basis of pairs AB, BC, CD et DA determine 'virtual' network ABCD.

Co-occurrence Threads (2)



Interpreting Threads

- Two characteristics determine their interpretations:
- - *Connexity* of nodes: there is a path connecting any given pair of words
- - *Non-transitivity* of co-occurrences: no co-occurrence is guaranteed beyond binary relations

Several networks 'living together' in the text

29 réseaux à 2 éléments :

- *services, publics* (50 occ.)
- *aujourd, hui* (32 occ.)
- *tous, toutes* (24 occ.)
- *service, public* (24 occ.)
- *mais, aussi* (22 occ.)
- *j, ai* (22 occ.)
- *pouvoir, achat* (22 occ.)
- *philippe, villiers* (18 occ.)
- *dominique, voynet* (18 occ.)
- *gérard, schivardi* (18 occ.)
- *ma, candidature* (14 occ.)
- *toujours, plus* (14 occ.)
- *son, ses* (14 occ.)
- *%, pib* (14 occ.)
- *emploi, formation* (14 occ.)
- *mettre, place* (12 occ.)
- *ségolène, royal* (12 occ.)
- *parti, travailleurs* (12 occ.)
- *tout, peut* (12 occ.)
- *petites, retraites* (10 occ.)
- *langues, régionales* (10 occ.)
- *madame, monsieur* (10 occ.)
- *mes, chers* (10 occ.)
- *personnes, âgées* (10 occ.)
- *énergies, renouvelables* (10 occ.)
- *jean, marie* (8 occ.)
- *abrogation, lois* (8 occ.)
- *1, million* (8 occ.)
- *on, ferme* (6 occ.)

9 réseaux à 3 éléments :

- *il, y, a* (51 occ.)
- *au, premier, tour* (32 occ.)
- *sécurité, sociale, démocratique* (32 occ.)
- *engagé, e, s* (30 occ.)
- *changer, conditions, vie* (28 occ.)
- *augmentation, bas, salaires* (25 occ.)
- *impôts, taxe, sur* (17 occ.)
- *voter, extrême, droite* (17 occ.)
- *au, premier, tour* (32 occ.)
- *25, mai, 2005* (15 occ.)

4 réseaux à 4 éléments :

- *autre, est, possible, pourquoi* (52 occ.)
- *classes, populaires, dans, quartiers* (52 occ.)
- *nicolas, sarkozy, françois, bayrou* (44 occ.)
- *grandes, entreprises, profits, bénéfices* (38 occ.)

4 réseaux à 5 éléments :

- *22, avril, 2007, élection, présidentielle* (80 occ.)
- *minima, sociaux, logements, construire, europe* (46 occ.)
- *éducation, nationale, enfants, accès, culture* (44 occ.)
- *moins, euros, milliards, 300, net* (39 occ.)

2 réseaux à 6 éléments :

- *ils, ne, n, pas, ont, peuvent* (183 occ.)
- *depuis, 5, ans, quatre, vingt, ans* (94 occ.)

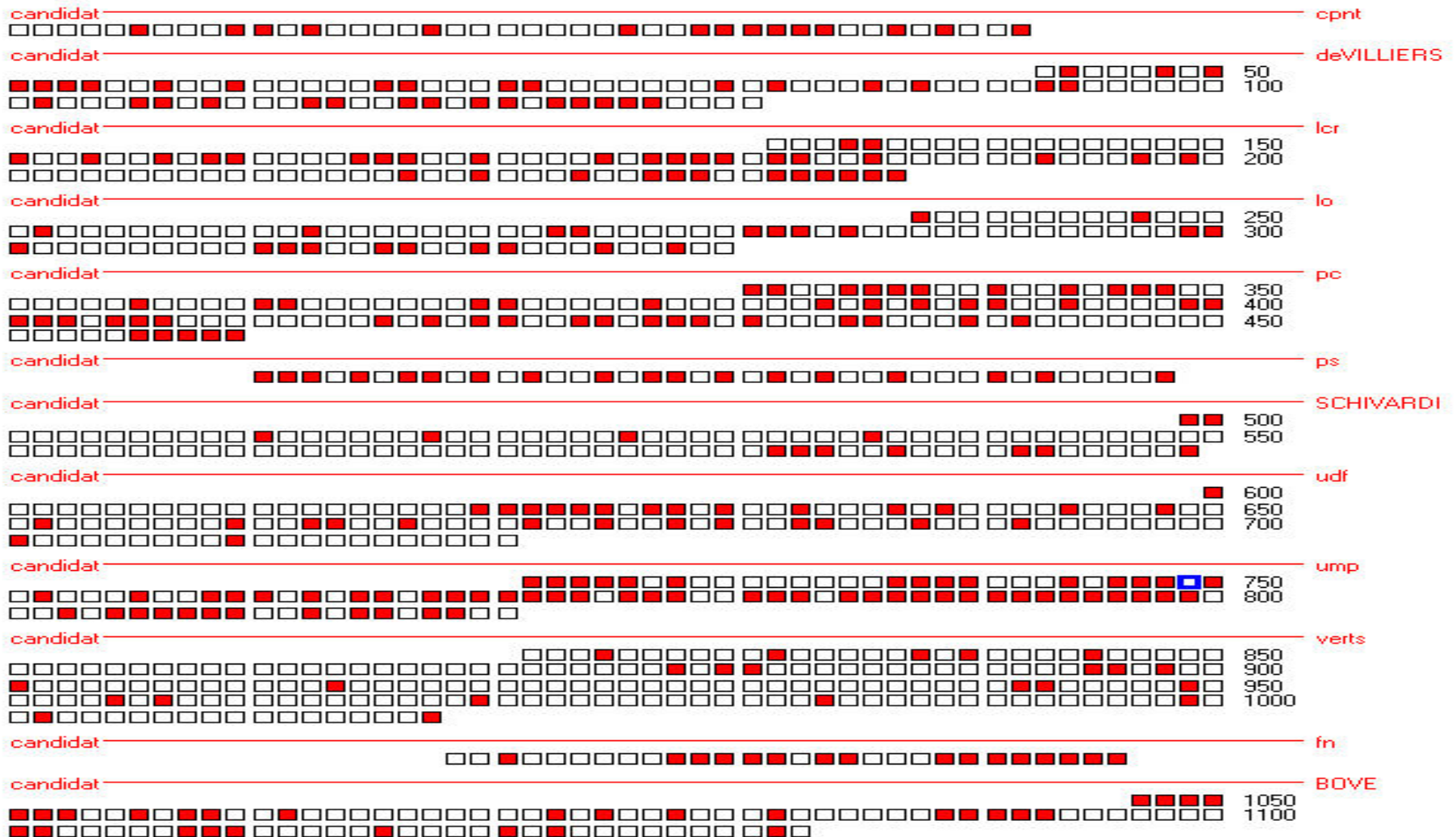
Ranking co-occurrence phenomena

■ *Lexical Valency* (magnetic power)

Forme	NbC	C/F	Forme	NbC	C/F	Forme	NbC	C/F
<i>nous</i>	7	0.07	<i>aux</i>	3	0.03	<i>sur</i>	2	0.02
<i>je</i>	6	0.04	<i>nos</i>	3	0.05	<i>avec</i>	2	0.02
<i>violences</i>	6	0.75	<i>entreprises</i>	3	0.08	<i>ne</i>	2	0.02
<i>contre</i>	5	0.10	<i>européenne</i>	3	0.09	<i>pas</i>	2	0.02
<i>être</i>	5	0.13	<i>veux</i>	3	0.11	<i>notre</i>	2	0.03
<i>femmes</i>	5	0.21	<i>depuis</i>	3	0.12	<i>a</i>	2	0.03
<i>faites</i>	5	0.71	<i>loi</i>	3	0.12	<i>s</i>	2	0.03
<i>vous</i>	4	0.05	<i>22</i>	3	0.13	<i>n</i>	2	0.03
<i>ans</i>	4	0.08	<i>président</i>	3	0.13	<i>pays</i>	2	0.04
<i>éducation</i>	4	0.18	<i>avril</i>	3	0.14	<i>vie</i>	2	0.04
<i>euros</i>	4	0.21	<i>2007</i>	3	0.15	<i>ont</i>	2	0.04
<i>présidentielle</i>	4	0.21	<i>suis</i>	3	0.30	<i>sociale</i>	2	0.05
<i>cadre</i>	4	0.36	<i>est</i>	2	0.10	<i>salaires</i>	2	0.07
<i>rupture</i>	4	0.36	<i>il</i>	2	0.02	<i>(...)</i>		

Ranking co-occurrence phenomena (2)

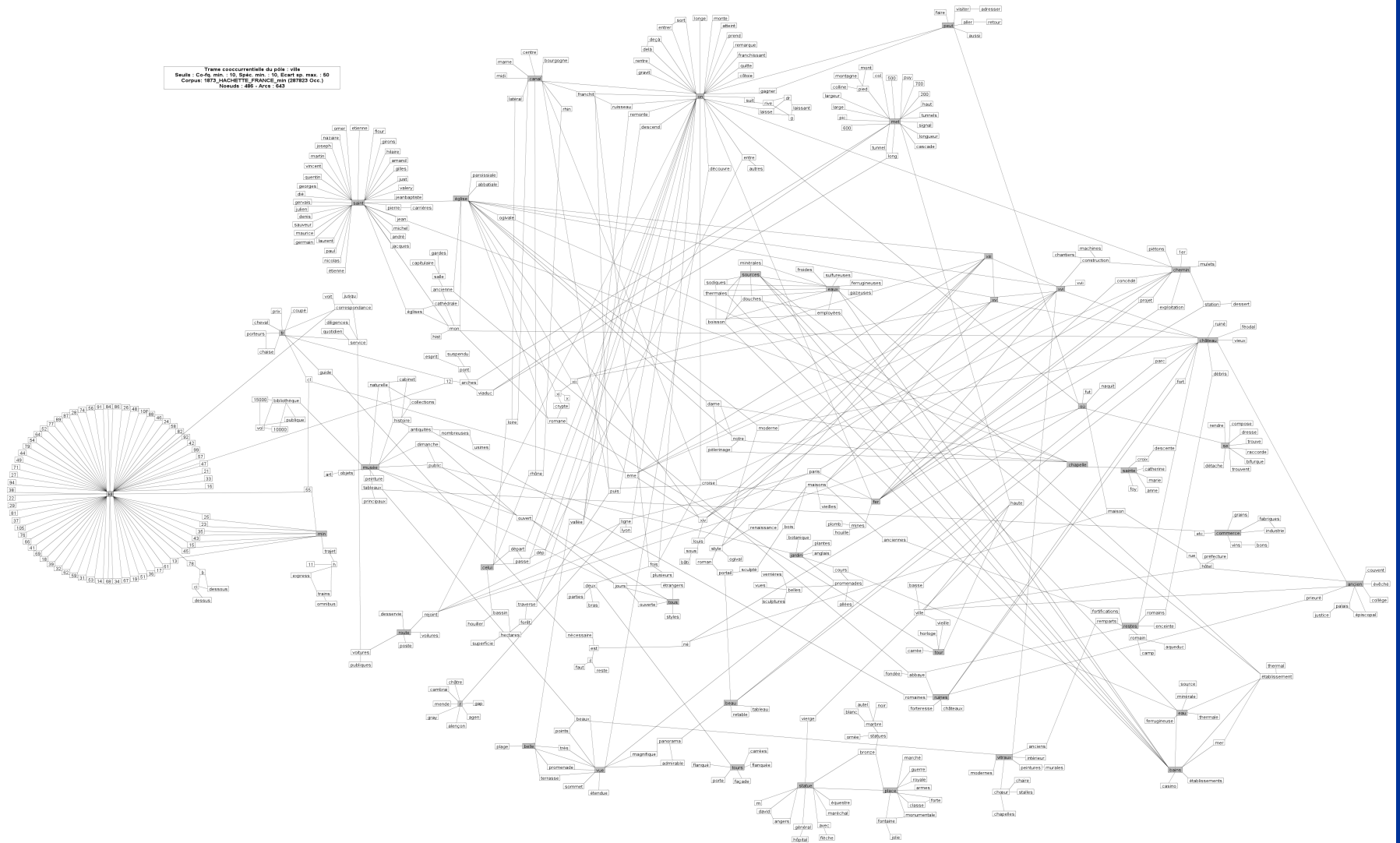
■ Co-occurrence surface



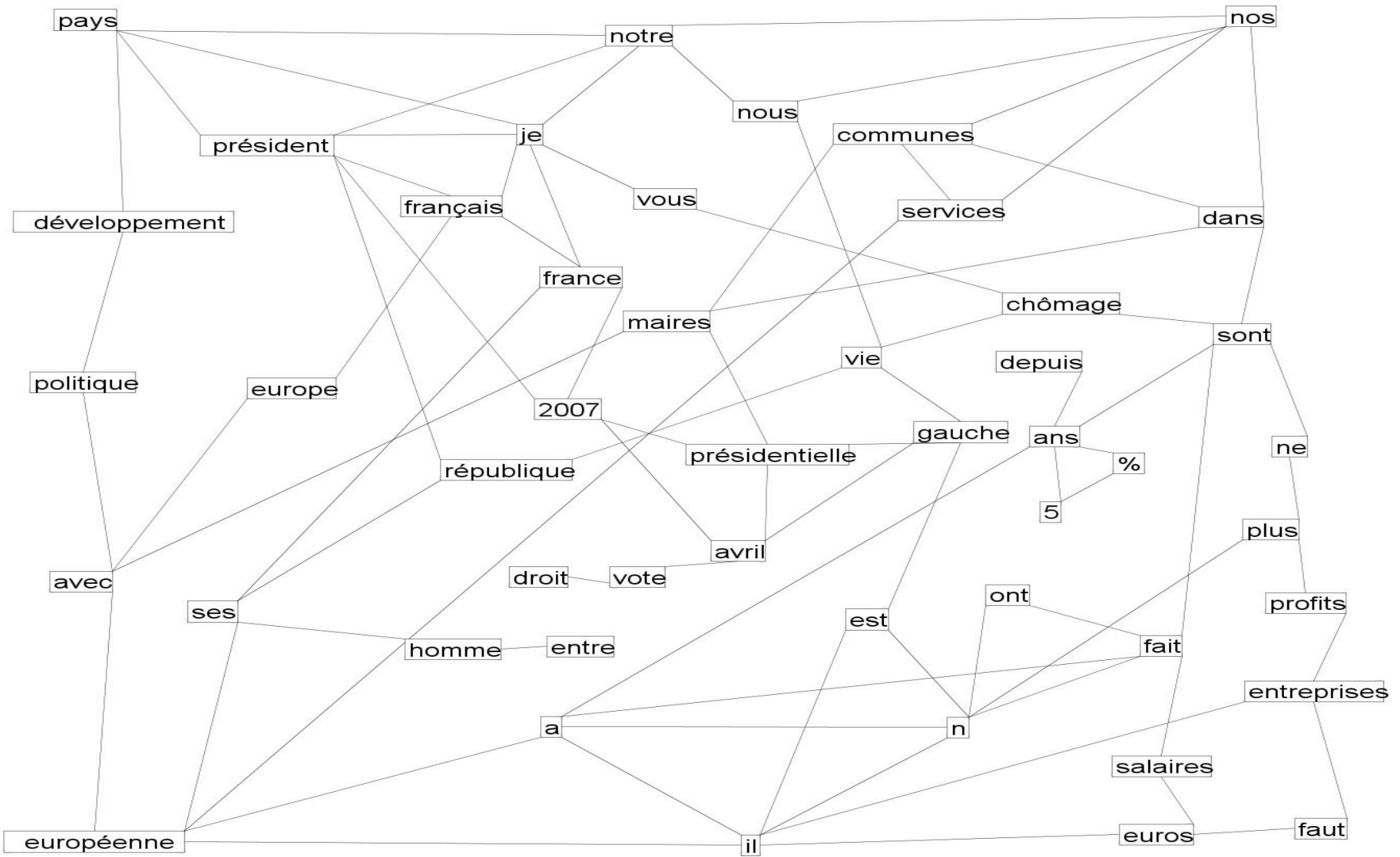
Hyper-Threads

- Deeper contextual exploration
- Determining concepts
- From information to knowledge?
- Another level of inter-corpus comparison

Thresholds and exploration levels

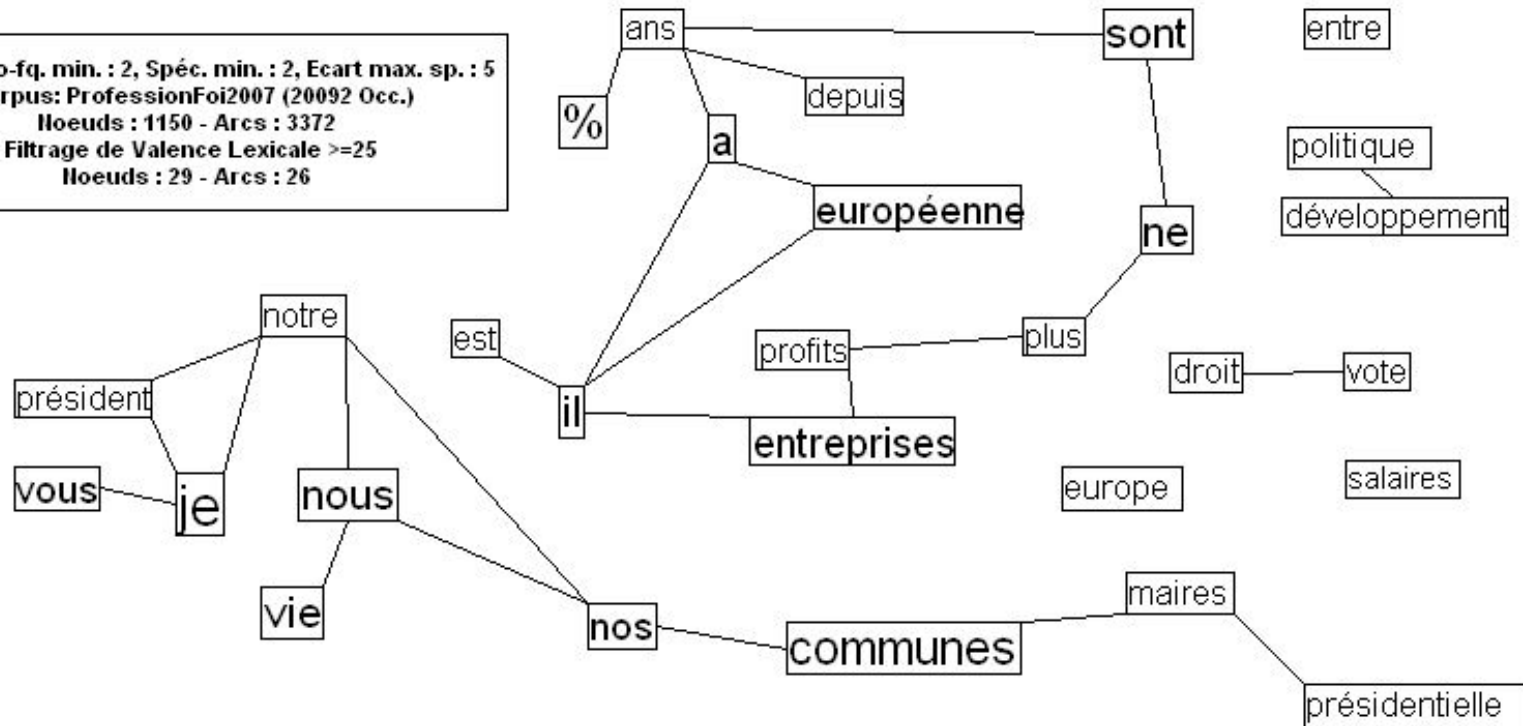


Hyper-Threads (2)



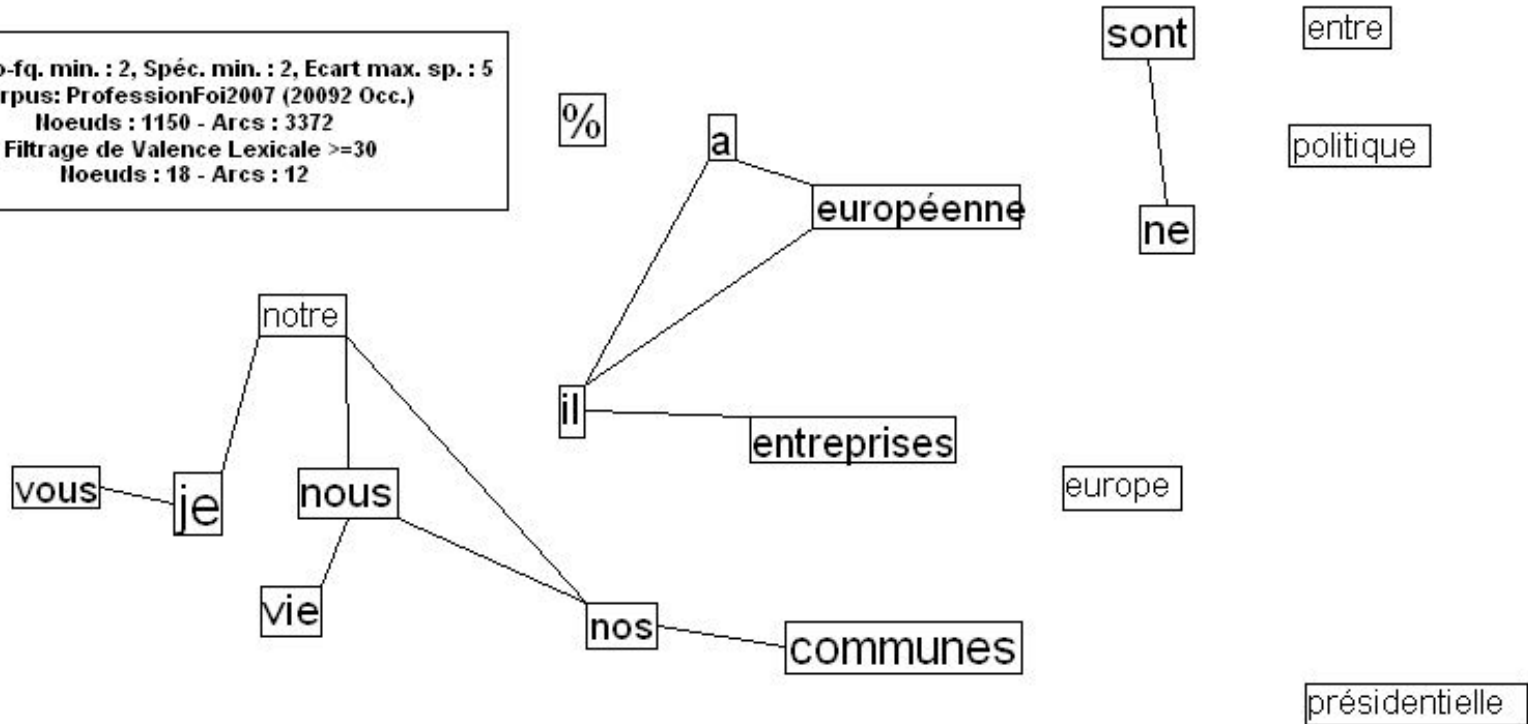
Hyper-Threads (3)

Seuils : Co-fq. min. : 2, Spéc. min. : 2, Ecart max. sp. : 5
Corpus: ProfessionFoi2007 (20092 Occ.)
Noeuds : 1150 - Arcs : 3372
Filtrage de Valence Lexicale >=25
Noeuds : 29 - Arcs : 26



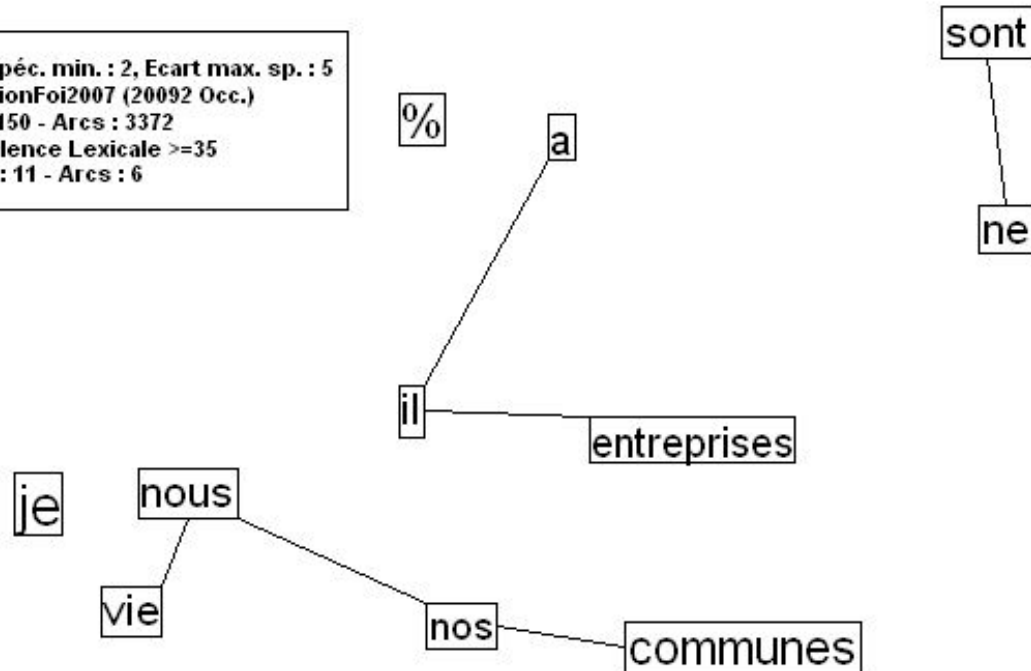
Hyper-Threads (4)

Seuils : Co-fq. min. : 2, Spéc. min. : 2, Ecart max. sp. : 5
Corpus: ProfessionFoi2007 (20092 Occ.)
Noeuds : 1150 - Arcs : 3372
Filtrage de Valence Lexicale >=30
Noeuds : 18 - Arcs : 12



Hyper-Threads (5)

Seuils : Co-fq. min. : 2, Spéc. min. : 2, Ecart max. sp. : 5
Corpus: ProfessionFoi2007 (20092 Occ.)
Noeuds : 1150 - Arcs : 3372
Filtrage de Valence Lexicale >=35
Noeuds : 11 - Arcs : 6



Hyper-Threads (6)

Seuils : Co-fq. min. : 2, Spéc. min. : 2, Ecart max. sp. : 5
Corpus: ProfessionFoi2007 (20092 Occ.)
Hoeuds : 1150 - Arcs : 3372
Filtrage de Valence Lexicale >=40
Hoeuds : 7 - Arcs : 1

sont

%

il

je

nous

vie

communes

Hyper-Threads (7)

Seuils : Co-fq. min. : 2, Spéc. min. : 2, Ecart max. sp. : 5
Corpus: ProfessionFoi2007 (20092 Occ.)
Moeds : 1150 - Arcs : 3372
Filtrage de Valence Lexicale >=45
Moeds : 1 - Arcs : 0

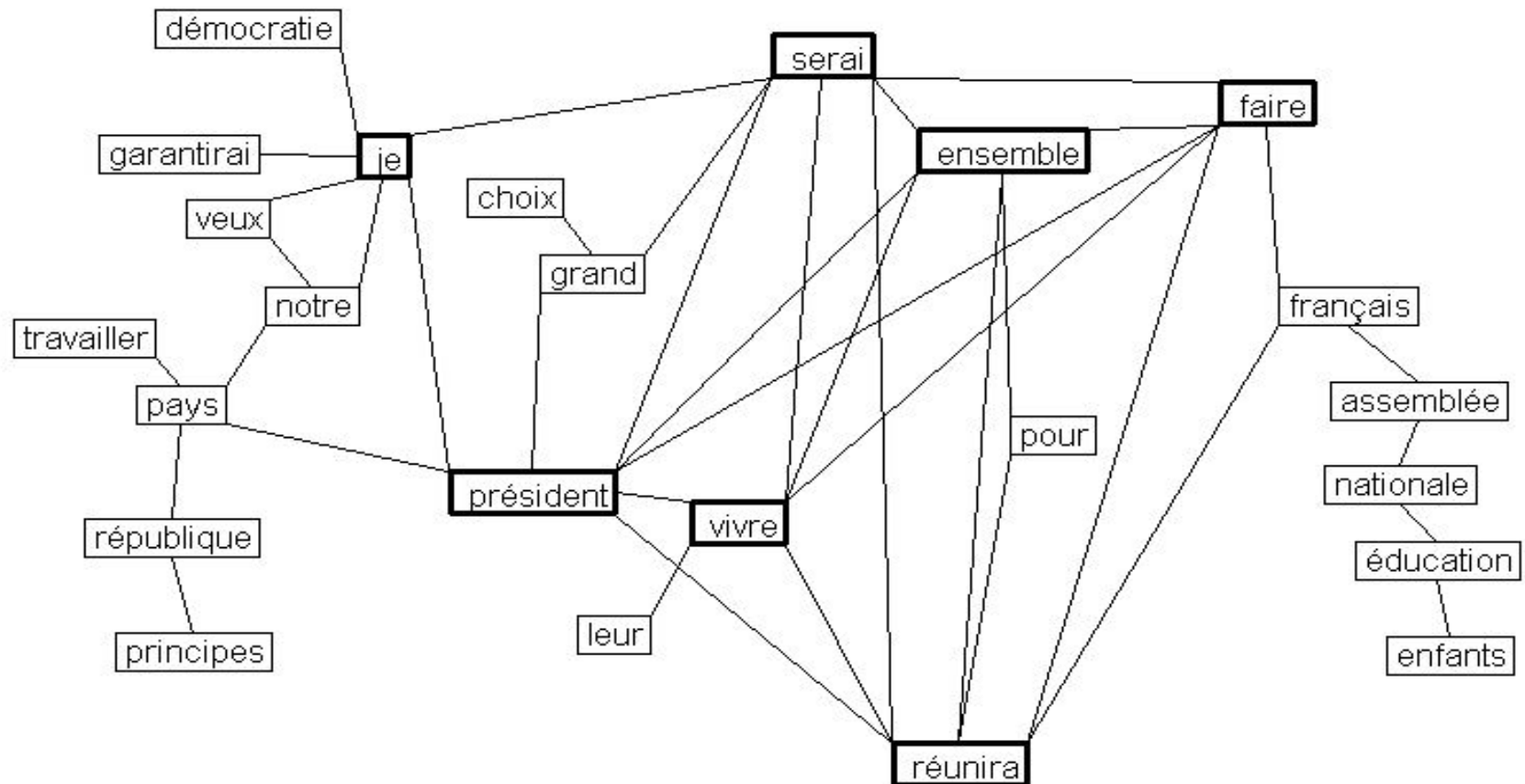
je

Le *je* chez les candidats de droite

- De Villiers, 26 occ., rang 9 du dictionnaire
- Bayrou, 25 occ., rang 9
- Le Pen, 7 occ., rang 9
- Sarkozy, 63 occ. rang 2

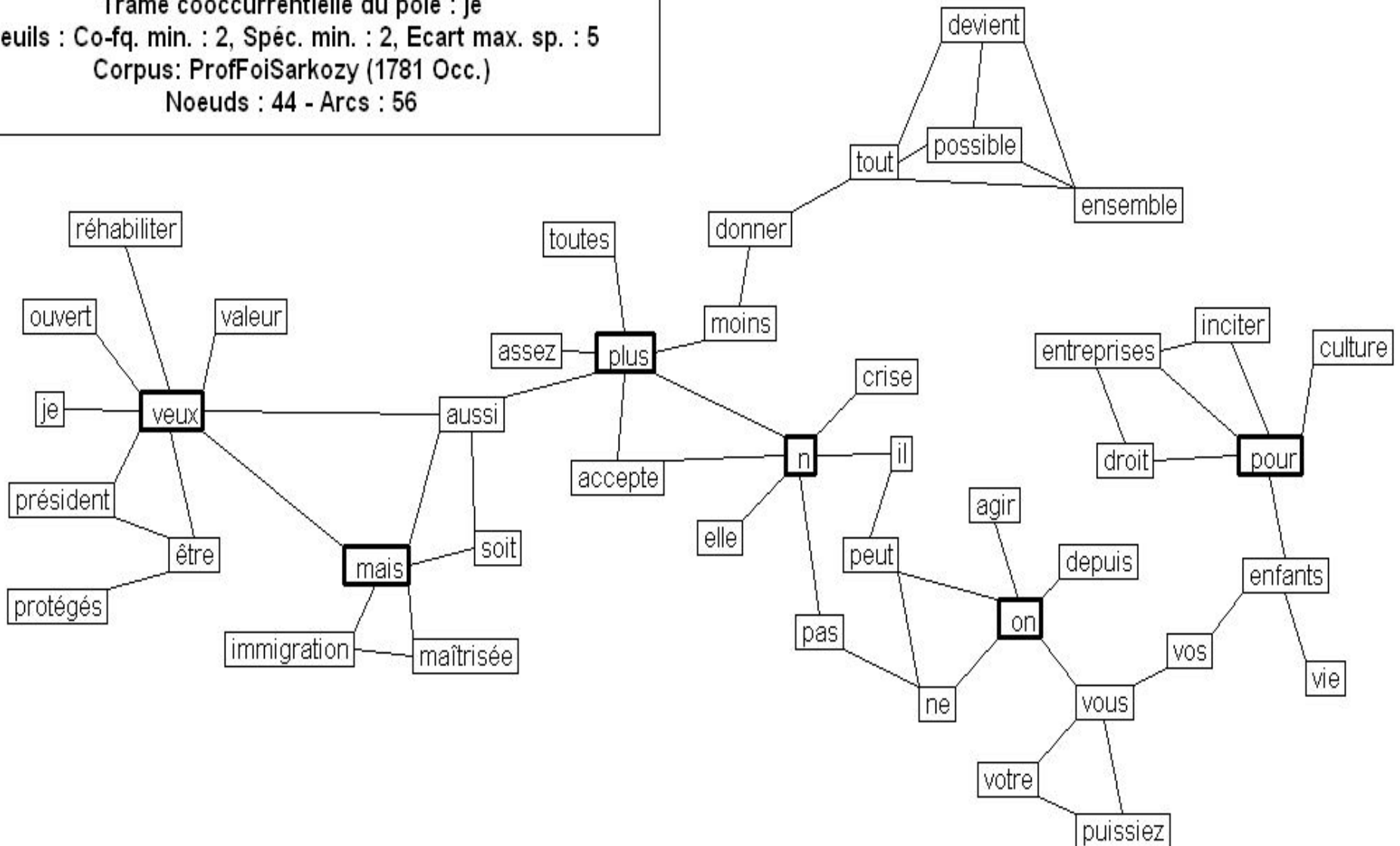
Trame : *je* chez Bayrou

Trame cooccurentielle du pôle : *je*
Seuils : Co-fq. min. : 2, Spéc. min. : 2, Ecart max. sp. : 5
Corpus: ProfFoiBayrou (1414 Occ.)
Noeuds : 24 - Arcs : 39



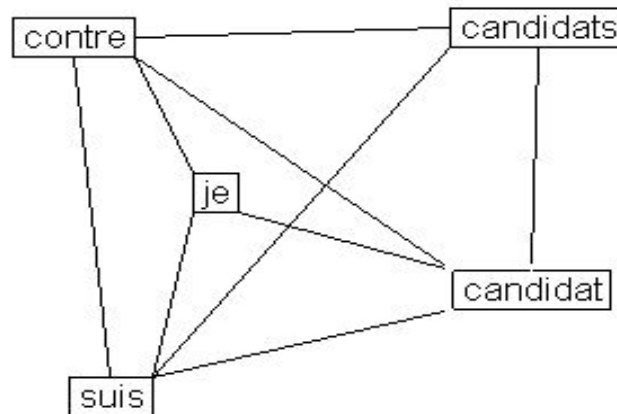
Trame : *je* chez Sarkozy

Trame cooccurentielle du pôle : je
Seuils : Co-fq. min. : 2, Spéc. min. : 2, Ecart max. sp. : 5
Corpus: ProfFoiSarkozy (1781 Occ.)
Noeuds : 44 - Arcs : 56



Trame : *je* chez LePen

Trame cooccurentielle du pôle : *je*
Seuils : Co-fq. min. : 2, Spéc. min. : 2, Ecart max. sp. : 5
Corpus: ProfFoiLePen (583 Occ.)
Noeuds : 5 - Arcs : 9



Remarks on terminology

- *text*

from Ancient Greek *textus*: to weave, to intertwine, to plait

- *analysis*

To study an object by decomposing it, from complex to simple
from Ancient Greek *analuein* : ‘undo a weft’

- *synthesis*

To explain an object by recomposing, from simple to complex
from Ancient Greek *synthesis* : ‘to lay together’

Applications to Terminology

pôle		<i>problème</i>		<i>problèmes</i>	
		(F = 228)		(F = 208)	
cooccurrent	F	f	sp	f	sp
<i>douloureux</i>	13	6	+E8	non spéc.	
<i>complexe</i>	18	non spéc.		non spéc.	
<i>complexes</i>	13	non spéc.		4	+E5
<i>endettement</i>	21	4	+E4	non spéc.	
<i>voisinage</i>	16	non spéc.		6	+E7

ée au **douloureux problème** des enfants de couples mixtes
er ce **douloureux problème** de la bosnie. à la demande de la
as le **douloureux problème** du chômage, mais notre pays dispo
ur le **douloureux problème** des otages. enfin, chaque fois qu
nt de nouveau ce **problème douloureux** pour dire que les deux
derai d'abord le **problème** particulier et toujours **douloureux**

qui viennent des **problèmes complexes** à résoudre, notamment t
de résoudre des **problèmes complexes** que l'histoire n'a pa
vra résoudre des **problèmes complexes**, très liés entre eux, de
l'ensemble des **problèmes** en tant ceux-ci sont **complexes**,

Applications to Terminology (2)

<auteur=raimond>

s'agissant des mesures prises par le gouvernement soviétique, sur lesquelles vous m'avez interrogé, nous n'avons pas, à ce stade, d'informations certaines - les soviétiques disposent de quinze réacteurs nucléaires du type de celui de tchernobyl.

<auteur=dumas>

j'ajouterai enfin que le comportement du gouvernement de la france n'a en aucune manière changé à l'égard du gouvernement libyen. comme vous le savez, des décisions ont été **prises** par les douze pays de la cee. elles ont été respectées et appliquées par le gouvernement de la république et si, à l'occasion de la reprise du dialogue euro-arabe au mois de décembre dernier à paris, la demande a été présentée à la cee par certains pays arabes de reconsidérer leur position à l'égard de la syrie comme à l'égard de la libye, la france attend, pour modifier son attitude, que des décisions nouvelles soient **prises** qui en changeraient le sens.

<auteur=dumas>

10 mf de mesures nouvelles pour l'europe de l'est qui voit ainsi confirmer son rang prioritaire, avec une dotation qui s'établira à un niveau quatre fois supérieur à celui de 1990.

