

Corpus annotés, enrichis, étiquetés

Les niveaux de l'analyse lexicométrique

Linguistique de corpus avril 2008

1. La catégorisation pour quoi faire?

- Outils lexicométriques, analyses contrastives ou chronologiques: Hyperbase, Lexico3
- Alceste, cooccurrenceur et constructeur de classes sémantiques
- Travail en surface (forme graphique) avec lemmatisation partielle par groupements de formes manuels (Lexico3) ou racinisation automatique (Alceste)
- Catégoriseurs et analyseurs de contenu (Cordial, Tropes)

Catégories morphosyntaxiques

Mes	DETPOSS	Ds1.ps
Chers	ADJMP	Afmp
Compatriotes	NCPIG	Nc.p
,	PCTFAIB	Ypw
\r\r		
Vous	PPER2P	Pp2.pn
m'	PPER1S	Pp1.sa
avez	VINDP2P	Vaip2p
élu	VPARPMS	Vmpasm
,	PCTFAIB	Ypw
en	PREP	Sp
mai	NCMIN	Ncm.
dernier	ADJMS	Afms
,	PCTFAIB	Ypw
pour	PREP	Sp
que	SUB	Cs
nous	PPER1P	Pp1.pn
construisions	VINDI1P	Vmii1p
ensemble	ADV	Rgp
une	DETIFS	Da-fs-i
nouvelle	ADJFS	Afufs
France	NPFS	Npfs



Lexico 3

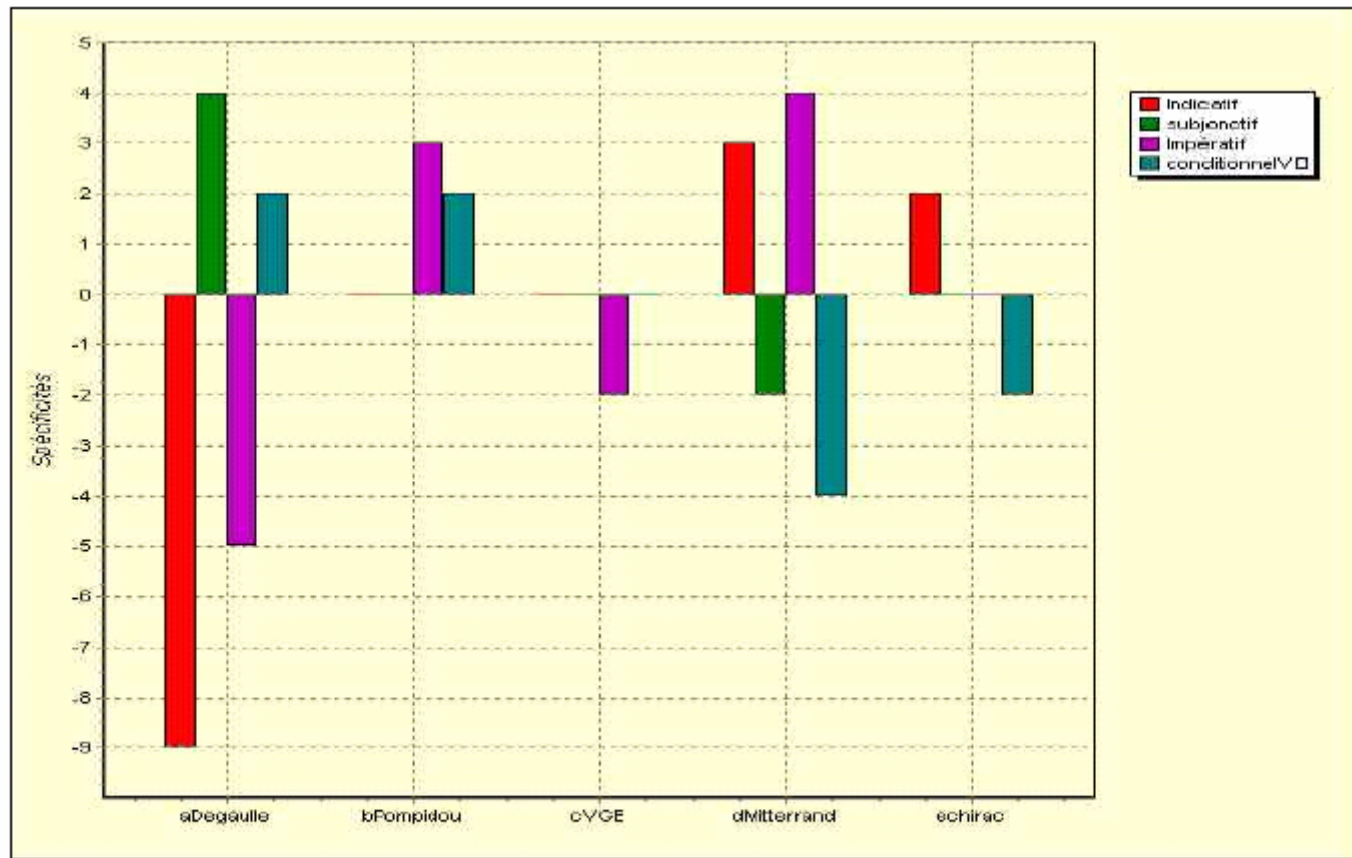
Morphosyntaxiques

+ VINF

Il faut
 Nous devons
 Au gouvernement de
 Pour
 Afin de

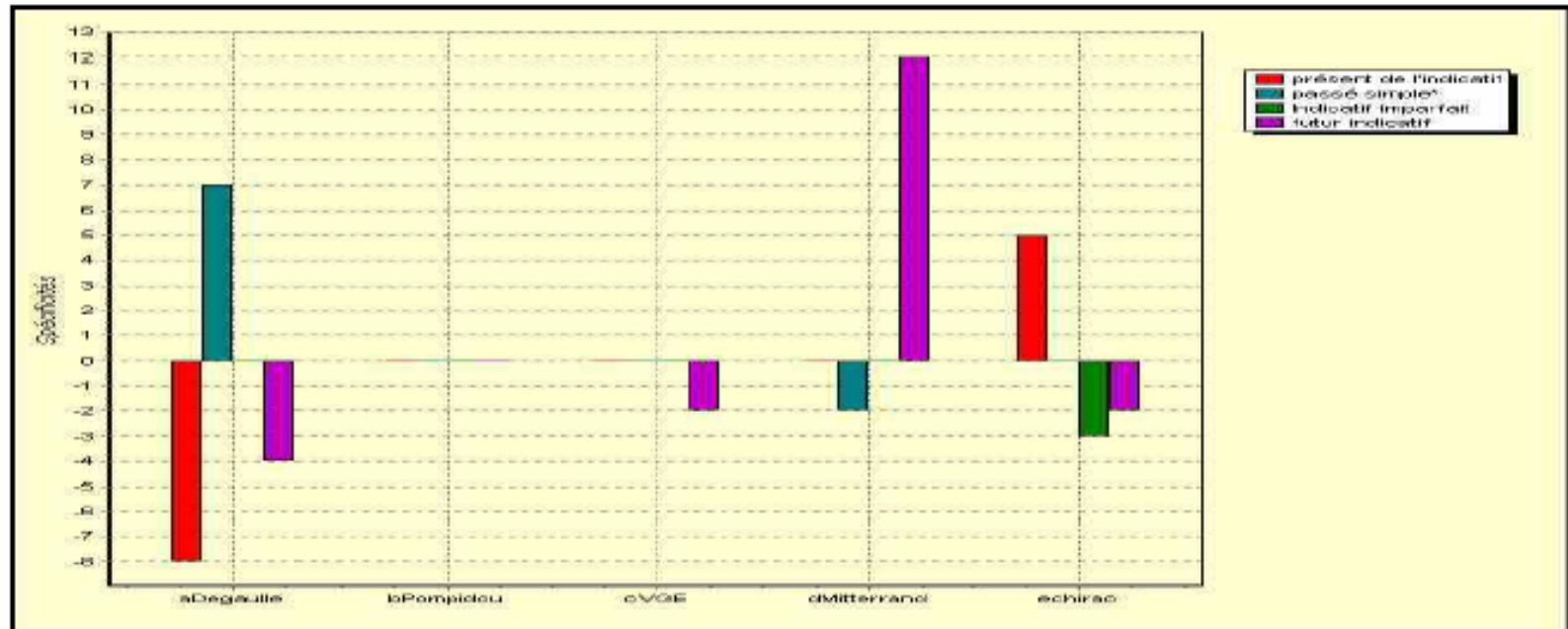
Forme	Fréq. Tot.	Coeff.	Forme	Frq. Tot.	Fréquence	Coeff.	
DETIFS	350	124	10	DETDFS	1928	375	-2
VINF	1330	366	8	PREP	5468	1101	-2
NCFP	899	244	5	ADJINT	30	2	-2
NCPIG	137	48	4	VIMPP2P	28	1	-2
NCFS	2852	677	4	VINDI3S	72	8	-2
PPER1P	654	166	3	VCONP3S	43	3	-2
ADV	2388	560	3	VINDF3S	210	31	-2
VINDP1P	311	87	3	NCMP	1027	197	-2
ADJFS	670	162	2	COO	1628	315	-2
VPARPFS	141	40	2	PRI	1017	191	-2
ADJFP	298	77	2	NPMS	93	9	-3
VINDP3P	542	132	2	SUB	897	145	-5
VINDP3S	1537	357	2	NCMIN	794	122	-5
NCFIN	106	13	-2	VPARPM S	497	71	-5
				ADJNUM	236	19	-8

Catégories morphosyntaxiques



Spécificités des modes sur le corpus vœux eti_1 (lexico3)

Les temps verbaux

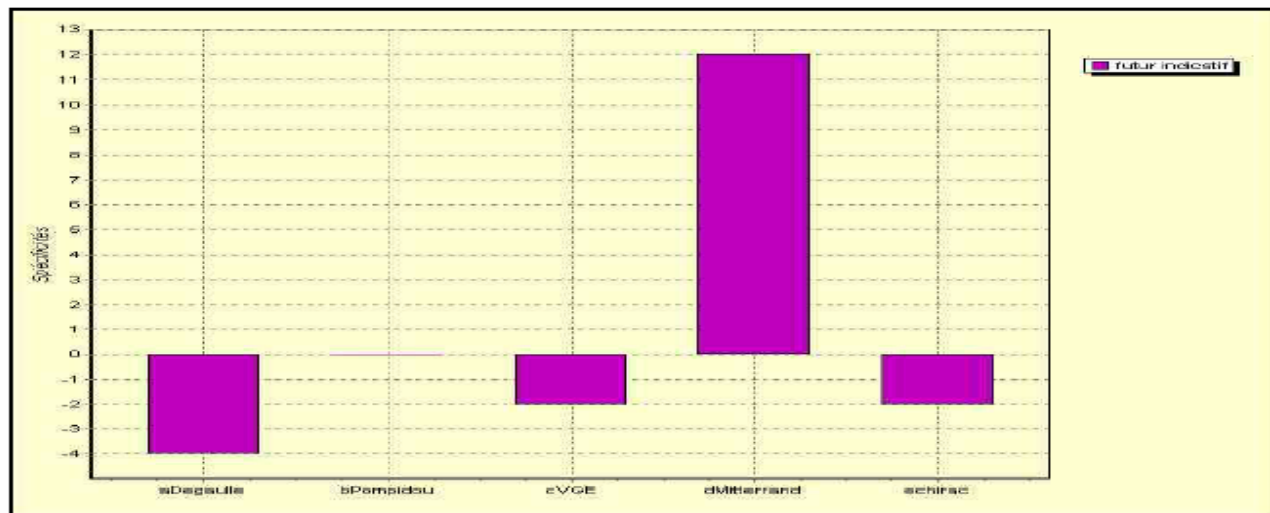


Histogramme des spécificités des temps de l'indicatif

Catégories morphosyntaxiques



Distribution du futur de l'indicatif en fréquences absolues



Histogramme des spécificités pour le futur de l'indicatif sur la partition locuteur

Catégories morphosyntaxiques

Formes (ordre lexicométrique)	Fréquence
	5481
	5468
	4462
	2864
	2615
	2197
	2157
	1872
	1628
	1450
	1303
	1266
	1199
	1153
	1034
	991
	899
	897
	852
	794
	739
	737
	610
	607
	595
	582
	572
	540
	539
	527
	497
	480
	445
	349
	334
	331
	220

<locuteur=aDegaulle>

Sp Da - fs - d Ncfs Afpfs , Sp Da - ms - d Npfs , Sp Da - fs - d Ncfs , Pp1 . sn Vmip1s Da - . p - i Ncm
Afpmp Cc Afpmp Da - ms - d Ao - . . Ncms Sp Ncm . . Pp1 . sn Vaip1s Vmpasm Sp Da - ms - d Ncms
- . . - Dd - fs - Ncfs Pp1 . pd Vmif3s Afp . s , Cs Pp1 . pn Vaip1p Vmpasm Rgp Da - ms - d Ncm . Sp P
fsa Pr - . . n Vmip3s . Sp Npfs Rgp , Ds1 . p . Ncfp Vmip3p Sp Da - ms - d Np . . Da - ms - d Ncfs Cc I
- ms - d Ncfs Pr - . . n Pp3 . sd Vmip3p Sp Vmn - . . Pp3msn Pp3 . . d Vmif3s Rgp , Rgp . Ds1 . p . Ncf
Vmip3p Sp Ncms Cc Pp3ms - Vmif3p Rgp . Da - ms - d Ncms Sp Ncfs Da - . p - i Ncfp , Sp Da - fs - d
Ncfs , Sp Da - fs - d Ncfs , Da - ms - d Ncms , Da - . p - i Ncm Cc Sp Da - fs - d Ncfs , Ds1 . s . Ncfs
Ds1 . s . Ncfs , Ds1 . s . Ncms , Vmip3p Sp Da - fs - i Afpfs Ncfs Sp Da - ms - d Ncfs Cc Da - ms - d
Ncm . Pr - . . n Vmip3p Vmn - . . Ncm . . Da - ms - d Ncms Afpms Vmip3s Da - ms - d Ncms Sp Dd - fs
Afpfs Ncfs . Sp Da - . p - d Ncm Afpms , Afpms , Afp . s , Rgp , Da - ms - d Ncms Vmif3s Pi - msa Sp
Cs Da - . p - d Ncfp Sp Rgp Vmip3p Da - fs - d Ncfs Cc Da - fs - d Ncfs Sp Da - . p - d Ncfp Afpfp , D
. p - d Ncfp , Da - . p - d Ncm Sp Da - fs - d Ncfs Afpfs . Sp Npfs , Da - . p - d Ncm Afpmp Vmip3p
Sp Ncfs Cc Sp Ncfs . Da - fs - d Ncfs , Sp Da - fs - d Ncf . Vaip3s Vmpasf . Pi - mpn Vmip3p Cc Vmip
Px3 . . a Pp3 . . Vmn - . . Cs Da - . p - d Ncm Vmsr3p Rgp Vmn - . . Sp Pp3mpd Sp Da - ms - d Ncms
Cs Dt - . s - Ncfs Vmsr3s Ncfs , Sp Dt - fs - Ncfs Cc Sp Dt - fs - Ncfs , Da - . p - d Ncm Pr - . . n
Vmif3p Dd - fs - Afpfs Ncfs , Cc Pr - . . - Vmsr3s Vmpasf Da - fs - d Ncfs Pr - . . n Vmif3s Sp Da - ms -
Npfs Da - ms - i Ncm . Afp . s Cc Afp . s . Sp Da - fs - d Ncfs , Da - . p - d Ncm Vmip3p Ds3 . sp Ncf
Sp Ncms Sp Da - fs - d Npfs Afpfs . Pd - fsn Vmip3s Vmn - . . Sp Vmn - . . Ds3 . ss Ncm . Sp Ds3 . pp
Afp . p Np . . Sp Cs Pp3mpn Vmip3p Ds3 . pp Ncfp Cc Vmip3p Ds3 . pp Afp . p Ncm . Sp Vmpp - . S
Pp3fsd Da - ms - i Afpms Rgp Sp Ncfs Afpfs , Sp Ncfs , Sp Ncfs , Sp Ncfs , Pr - . . n Vmsr3s Da - fs - d
Ncfs Afpfs Cc Da - fs - d Ncfs . Da - fs - d Npfs , Pr - . . - Da - fs - d Ncfs , Da - . p - d Ncfp , Da - fs -
Ncfs , Vmip3p Sp Afpms Ncms , Vmfé2s , Sp Ncms , Da - ms - d Ncms Afpms , Ds3 . ss Ncms Afpms .
Sp Ncm . Pp3fsn Vmip3s Vmn - . . Sp Da - fs - d Ncfs Da - . p - i Ncm Sp Da - ms - d Ncm . Cc Da - r
- d Ncm . , Cc Rgp , Da - ms - d Ncms Sp Da - fs - d Ncfs Sp Dd - . p - Ncm Rgp Vmpapm Sp Ncfs S
Vmn - . . Mc . p Ncm Sp Ncm Sp Vmn - . . Sp Ds3 . sp Afpms Da - fs - d Ncfs . Pp3fsn Vmip3s Vmn -
Da - . p - d Ncm Pr - . . n Pp3 . sa Vmip3p Sp Dt - . p - Np . . Sp Da - ms - d Npfs Afpfs , Sp Vmpp -
Cs Vmsr3s , Da - ms - i Ncms Px3 . . Vmn - . . Da - ms - d Npfs Rgp Afpfs Sp Da - ms - d Ncms Cc Sp
Da - fs - d Ncf . . Cc , Sp Dt - fs - Ncfs , Da - fs - d Npfs Vmip3s Rgp Vmn - . . Sp Da - fs - d Ncfs Cc ,
Da - fs - d Ncfs , Da - ms - d Ncms Sp Da - ms - d Ncfs Afp . s . Rgp , Da - ms - d Ncm . Da - . p - i

Sélectionnez une couleur : ■

cherche :

mes (ordre lexicométrique)	Fréquence
	5481
	5468
	4462
	2864
	2615
	2197
	2157
	1872
	1628
	1450
	1303
	1266
	1199
	1153
	1034
	991
	899
	897
	852
	794
	739
	737
	610
	607
	595
	582
	572
	540
	539
	527
	497
	480
	445
	349
	334

<locuteur=A.degaulle>

<texte=T1>

<année=1959>

\$ pour la métropole française , pour l ' algérie , pour la communauté , je forme des vœux ardents et confiants au premier jour de 1960 . je suis rempli de l ' espoir que cette année nous sera propice , parce que nous avons fait beaucoup au cours de celle qui finit .

\$ en france même , nos institutions assurent à l ' état l ' efficacité et l ' autorité qui lui permettent d ' agir . il en sera ainsi , désormais . nos finances sont en équilibre et le resteront demain . au point de vue des ressources , de la technique , de la recherche , du crédit , des échanges et de la monnaie , notre industrie , notre agriculture , notre commerce , disposent d ' une bonne base pour l ' expansion et le progrès qui vont marquer 1960 . le franc nouveau est le signe de cette féconde solidité . dans les domaines politique , social , scolaire , etc . , le pouvoir fera tout pour qu ' aux querelles d ' autrefois succèdent la concorde et la coopération entre les familles spirituelles , les catégories , les citoyens de la nation française .

\$ en algérie , les combats fratricides diminuent d ' ampleur et de fureur . la voie de la paix est tracée . tous peuvent et doivent s ' y rencontrer afin que les algériens puissent effectivement disposer d ' eux - mêmes par le suffrage , que chaque tendance prenne part , en toute liberté et en toute sécurité , aux débats qui prépareront cette grande consultation , et que soit hâtée la transformation qui fera de l ' algérie un pays moderne et prospère .

\$ dans la communauté , les peuples poursuivent leur évolution en accord avec la république française . celle - ci va continuer de prêter son concours à leurs jeunes états pour qu ' ils portent leurs responsabilités et développent leurs propres pays en formant avec elle un grand ensemble d ' activité économique , de culture , de défense , d ' influence , qui serve la condition humaine et la civilisation .

\$ la france , dont la population , les ressources , la puissance , sont en plein essor , joue , à présent , au plan voulu , son rôle international . en 1960 elle entend concourir à la détente des rapports entre l ' est et l ' ouest , et peut - être , au début de la coopération de ces peuples bien pourvus en vue d ' aider deux milliards d ' hommes à vaincre à leur tout la misère . elle entend resserrer les liens qui l ' unissent à d ' autres états de l ' europe occidentale , en attendant que puisse , un jour s ' unir l ' europe tout entière dans l ' équilibre et dans la paix . mais , à toute éventualité , la france entend également contribuer à la réforme et , par la suite , au

2. Weblex: catégorisation pour quoi faire?

- Un outil lexicométrique complet;
- Format d'entrée XML ou txt; tableaux
- Corpus textuels prétraités (segmentation; partitions)
- Ergonomie regroupée: réglages multiples des fonctions
- Documentation développée
- Outil de navigation textométrique (langage CQP)
- Travail en surface (forme graphique) ou sur des données catégorisées
- Une fonction originale: les lexicogrammes

Autour du JE présidentiel

cooccurents gauches

	f	cf	p	d _m
<u>chers</u>	55	21	4e-06	7.3
<u>métropole</u>	18	10	3e-05	11.1
<u>com patriotes</u>	62	20	1e-04	7.2
<u>françaises</u>	41	15	2e-04	10.7
<u>outre-mer</u>	14	7	1e-03	8.6
<u>voeux</u>	81	21	2e-03	4.9
<u>oeur</u>	27	9	8e-03	4.3
<u>vois</u>	6	3	4e-02	17.0

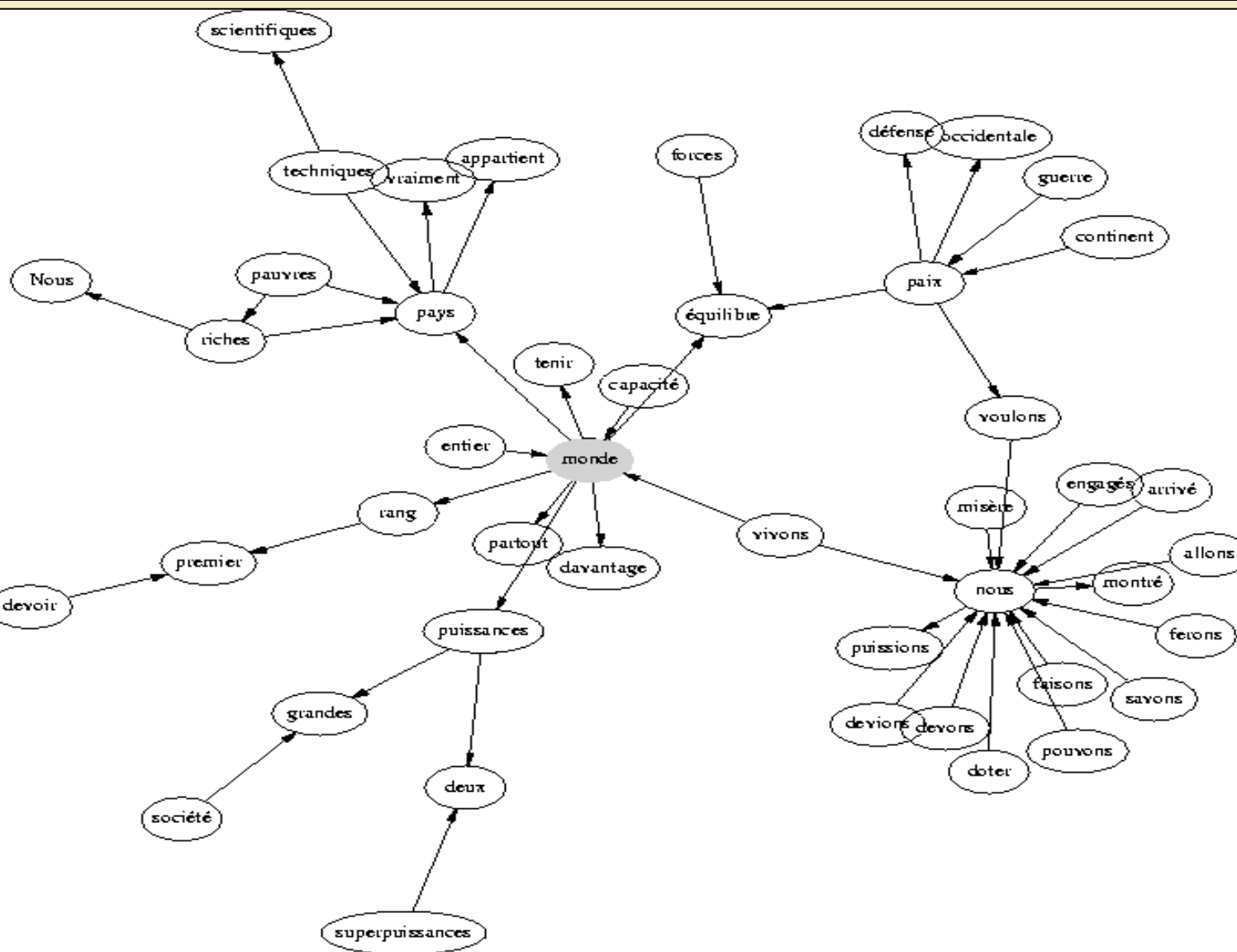
je (346) cooccurents droits

	f	cf	p	d _m
<u>souhaite</u>	65	64	6e-57	0.2
<u>pense</u>	22	21	9e-18	0.0
<u>voudrais</u>	19	19	3e-17	0.1
<u>sais</u>	19	18	4e-15	0.6
<u>forme</u>	19	16	8e-12	0.0
<u>crois</u>	11	11	3e-10	0.1
<u>veux</u>	11	11	3e-10	0.1
<u>dis</u>	12	11	3e-09	0.7
<u>dire</u>	39	19	1e-07	3.3
<u>adresser</u>	7	7	9e-07	6.9
<u>fais</u>	6	6	7e-06	0.8
<u>heureuse</u>	22	12	7e-06	13.5
<u>voeux</u>	81	26	9e-06	7.3
<u>soir</u>	42	17	1e-05	6.8
<u>bonne</u>	75	24	2e-05	8.5
<u>demande</u>	7	6	4e-05	2.2
<u>disais</u>	5	5	5e-05	0.8
<u>vois</u>	6	5	3e-04	0.0
<u>vais</u>	4	4	4e-04	0.0
<u>année</u>	202	42	7e-04	10.1
<u>adresse</u>	17	8	9e-04	1.0
<u>fier</u>	5	4	2e-03	6.5
<u>faibles</u>	5	4	2e-03	20.8
<u>salue</u>	3	3	3e-03	0.7
<u>dirai</u>	3	3	3e-03	1.0
<u>répète</u>	3	3	3e-03	1.3
<u>souviens</u>	3	3	3e-03	1.3
<u>assure</u>	3	3	3e-03	14.0
<u>sûr</u>	12	6	3e-03	1.0
<u>rendre</u>	9	5	4e-03	7.6
<u>promis</u>	6	4	4e-03	2.0
<u>espère</u>	6	4	4e-03	5.8
<u>propose</u>	6	4	4e-03	6.5
<u>dit</u>	14	6	8e-03	3.8
<u>fraternelle</u>	7	4	9e-03	17.8
<u>demandé</u>	4	3	9e-03	2.0
<u>satisfaction</u>	4	3	9e-03	12.0
<u>tire</u>	4	3	9e-03	19.3
<u>nom</u>	30	9	2e-02	6.4
<u>vivre</u>	30	9	2e-02	13.1
<u>parle</u>	8	4	2e-02	2.2
<u>constate</u>	5	3	2e-02	0.0
<u>ardents</u>	5	3	2e-02	3.0

**Lexicogramme simple
associé à une forme (je)
WEBLEX**

« Monde » dans les messages de vœux

1. Lexicogramme récursif de la forme « monde » sur l'ensemble du corpus

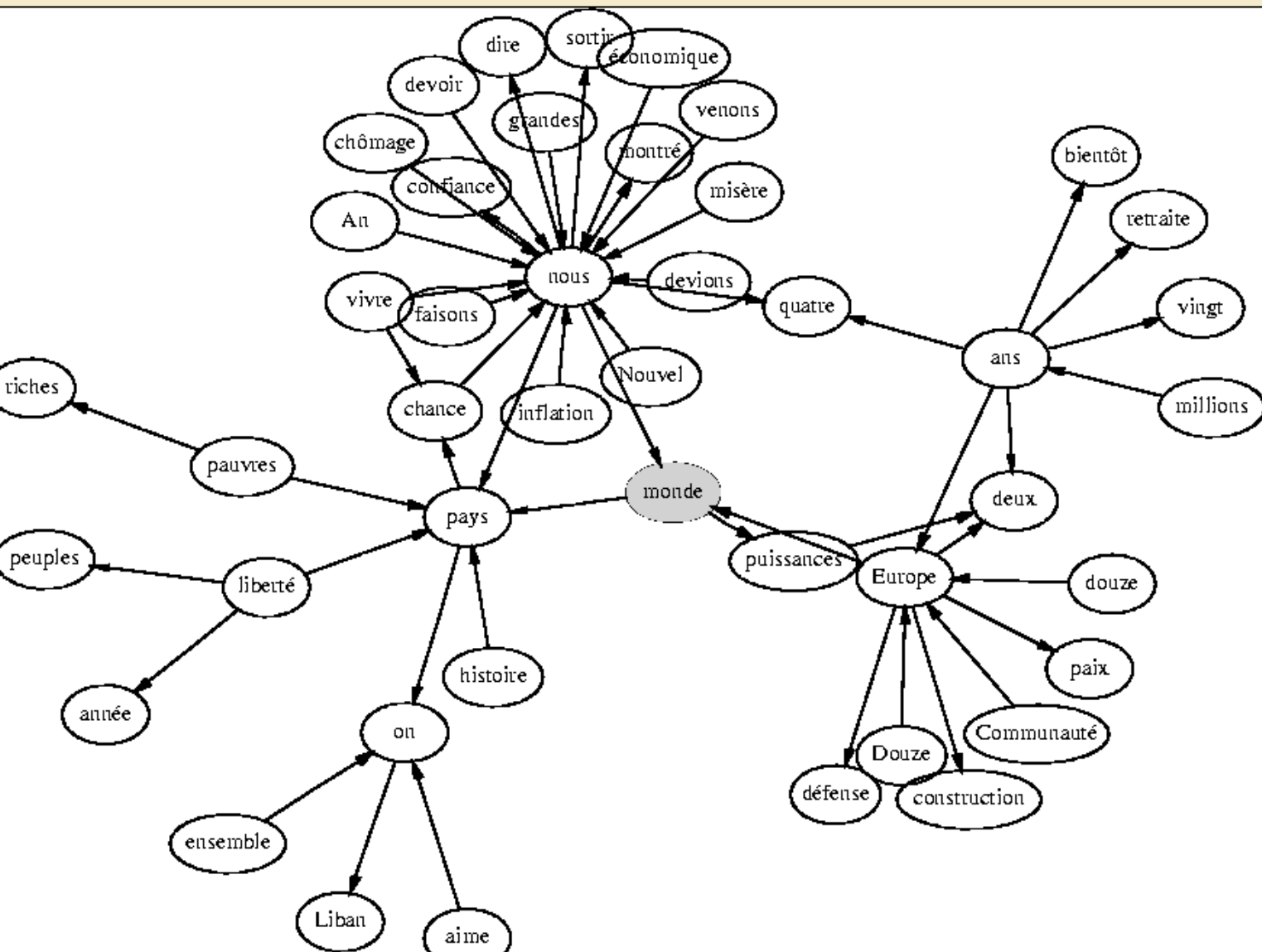


• Élagage des formes outils, numéraires et hapax.

• Seuils : f 3, cf 3, p 5.0E-2, dm 1000.0

« Monde » dans les messages de vœux

4. Lexicogramme récursif de la forme « monde » chez Mitterrand



- Élagage des formes outils numéraires et hapax.
- Seuils : f 3, cf 3, p 5.0E-2 dm 1000.0



