

Séminaire RUSS-INED - Paris, 21 janvier 2016

Exemple de base Biblio et d'utilisation de Macro Biblio

Annie Morin
annirisa@orange.fr

6 Exemple : base Biblio

Données et objectifs

- Données sur le lupus, résumés d'articles issus de PubMed , obtenus à partir d'une requête avec deux mots-clés : « lupus » et « clinical trials »
- Objectif : étudier les publications sur ce sujet depuis 1994 avec un intérêt particulier pour les médicaments proposés sur cette période.

Récupération des données

- Se connecter sur Pubmed :
<http://www.ncbi.nlm.nih.gov/pubmed>
- Requête avec deux mots-clés :
« lupus » et « clinical trials »
- Importation des données sous le format de votre choix.
- Les champs sont initiés par des tags
- On n'en conserve que quelques-uns :
titre, année, résumé, auteurs,
journal etc...

6. Exemple : base Biblio

Exemple (partie d'un enregistrement au format Medline)

PMID- 25084

DP - 2014 Aug 1

TI - Racial differences in treatment preferences among lupus patients:
a two-site study.

AB - OBJECTIVES: To identify the demographic, clinical and psychosocial characteristics associated with racial differences in willingness to receive cyclophosphamide (CYC) or participate in a research clinical trial(RCT) among
.....;

AU - Vina ER

AD - University

Traitement par TextomineR

- **Utilisation de la fonction MacroBiblio.**
 - `res.lupus<-MacroBiblio(lupus, lmd = 8, lmk=8)`
 - `print(res.lupus)`
 - `summary(res.lupus, nword=20)`
- **Original :**
 - `MacroBiblio(base, num.text = "Abstract",
num.agg = "Year", idiom = "en", lminword = 3,
Fmin = 10, Dmin = 5, Fmax = NULL, equivalence
= NULL, stop.word.user = NULL, lmd = 3, lmk =
3, ncp = 10, row.sup = NULL, col.sup = NULL,
graph = TRUE, axes = c(1, 2), proba = 0.01)`

Sorties (partielles)de MacroBiblio

- CA du tableau documents x mots
- CA du tableau agrégé par année
- Classification des années sous contraintes de contiguïté
- Analyse restreinte aux noms de médicaments et années

6. Exemple : base Biblio

CA du tableau documents x mots

- Number of documents = 506
- Corpus size = 91146
- Vocabulary size = 6244
- Frequency of some words :

	Frequency	N.Documents
the	4792	497
sle	2126	466
lupus	832	471
Systemic	588	453
erythematosus	542	448

6. Exemple : base Biblio

CA du tableau documents x mots

- Valeurs propres

	dim 1	dim 2	dim 3	dim 4
Variance	0.30912	0.27661	0.22788	0.22070
% of var.	1.49417	1.33703	1.10149	1.06679
Cum % of var	.1.49417	2.83120	3.93268	4.99948

6. Exemple : base Biblio

- Les 10 mots les plus fréquents après sélection

patients	2273	462
sle	2126	466
disease	864	310
lupus	832	471
activity	592	224
systemic	588	453
erythematosus	542	448
treatment	505	209
study	501	307
group	477	160

6. Exemple : base Biblio

Métaclés et métadocs dim1

- Metaclés + "association" "allele" "drb" "hla"
"gene" "susceptibility" "polymorphism"
"polymorphisms" "associated" "alleles"
"snps" "genetic" "risk" "controls" "genes"
"genotypes" "sle" "expression"
- Metadocs+` "373" "186" "250"
- Metaclés - "placebo" "treatment" "group"
"bmd" "prasterone« "months" "day"
"dhea" "dose"
- Metadocs- "151"

Métaclés et métadocs dim2

- Metaclés + "damage" "health" "disease"
"physical" "sdi" "hrqol" "social" "factors"
- Metadocs "350" "172" "441" "115"
- Metaclés "cells" "cell" "expression"
"depletion" "beta" "lymphocytes" "gene"
- Metadocs-

6. Exemple : base Biblio

Métaclés et métadocs dim3

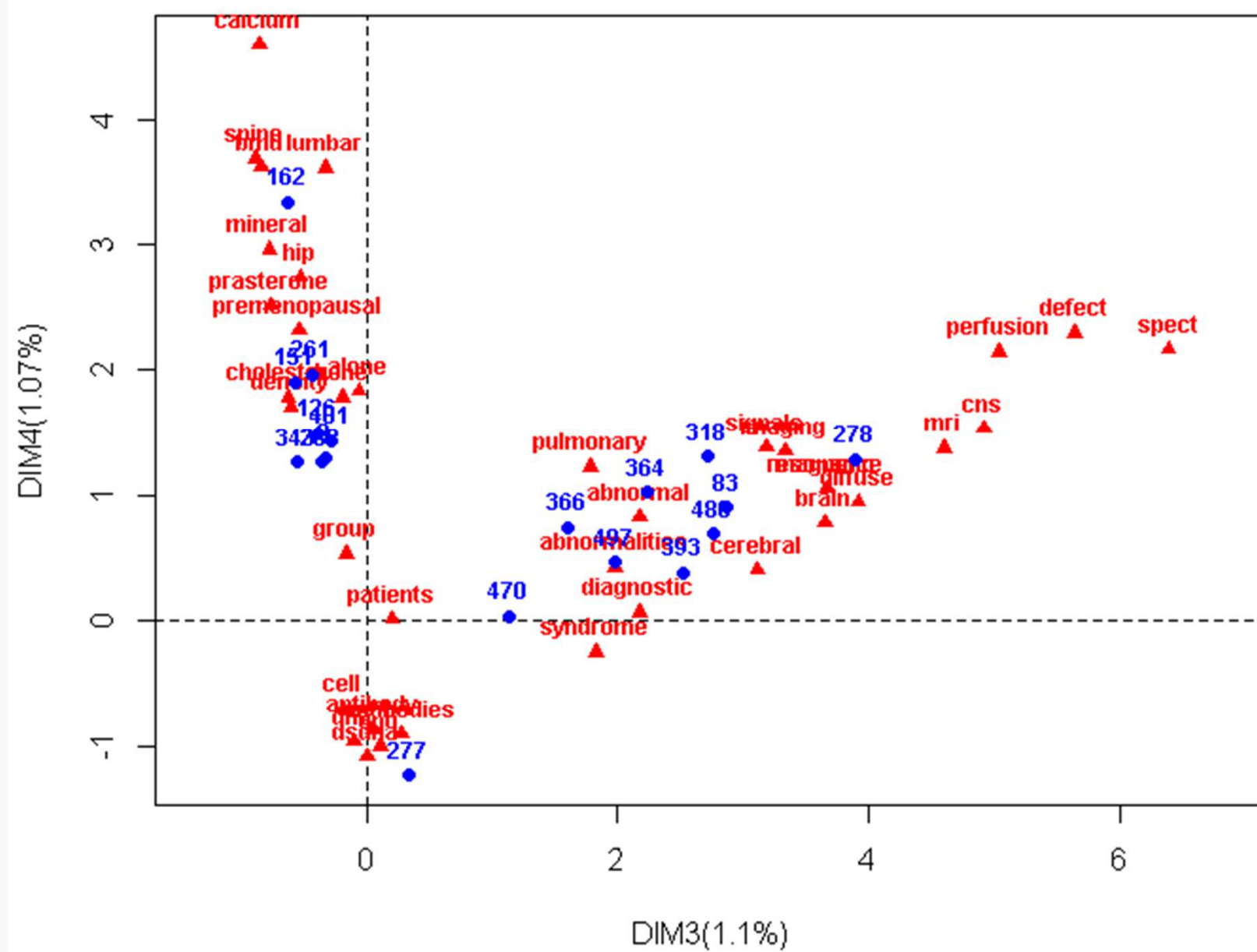
- Metaclés + "spect" "cns" "perfusion" "mri"
"defect" "brain" "imaging" "syndrome"
"diffuse" "abnormalities" "abnormal"
"magnetic" "resonance" "diagnostic"
"signals" "cerebral" "pulmonary" "patients"
- Metadocs+` "278" "83" "480" "318" "364"
"470" "497" "366" "393"
- Metaclés -
- Metadocs-

6. Exemple : base Biblio

Métaclés et métadocs dim4

- Metaclés + "calcium" "bmd" "spine"
"prasterone" "lumbar" "hip" "bone"
"cholesterol" "spect" "group" "density"
"perfusion" "mineral" "premenopausal" "alone"
- Metadocs + "162" "261" "151" "343" "126"
"401" "278" "8" "288"
- Metaclés - "anti" "dsdna" "antibodies"
"antibody" "dna" "cell"
- Metadocs- "277"

Metakeys & Metadocs



6-1. Exemple : base Biblio

Dimension des Métaclés

\$DimWord	Dim	Total	Dim	%Dim
bmd	6	10		60
spect	6	10		60
cns	5	10		50
drb	5	10		50
acl	4	10		40
calcium	4	10		40
cholesterol	4	10		40
costs	4	10		40
prasterone	4	10		40
pulmonary	4	10		40
spine	4	10		40
vascular	4	10		40
anti	3	10		30
cell	3	10		30
cells	3	10		30

6-1. Exemple : base Biblio

AC du tableau agrégé par année

- **Nombre de publications par année**

1994	1995	1996	1997	1998	1999	2000	2001	2002		
18	15	22	13	19	33	21	19	19		
2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	
19	30	38	35	42	34	31	39	34	25	

- **Valeurs propres de l'AC du tableau agrégé**

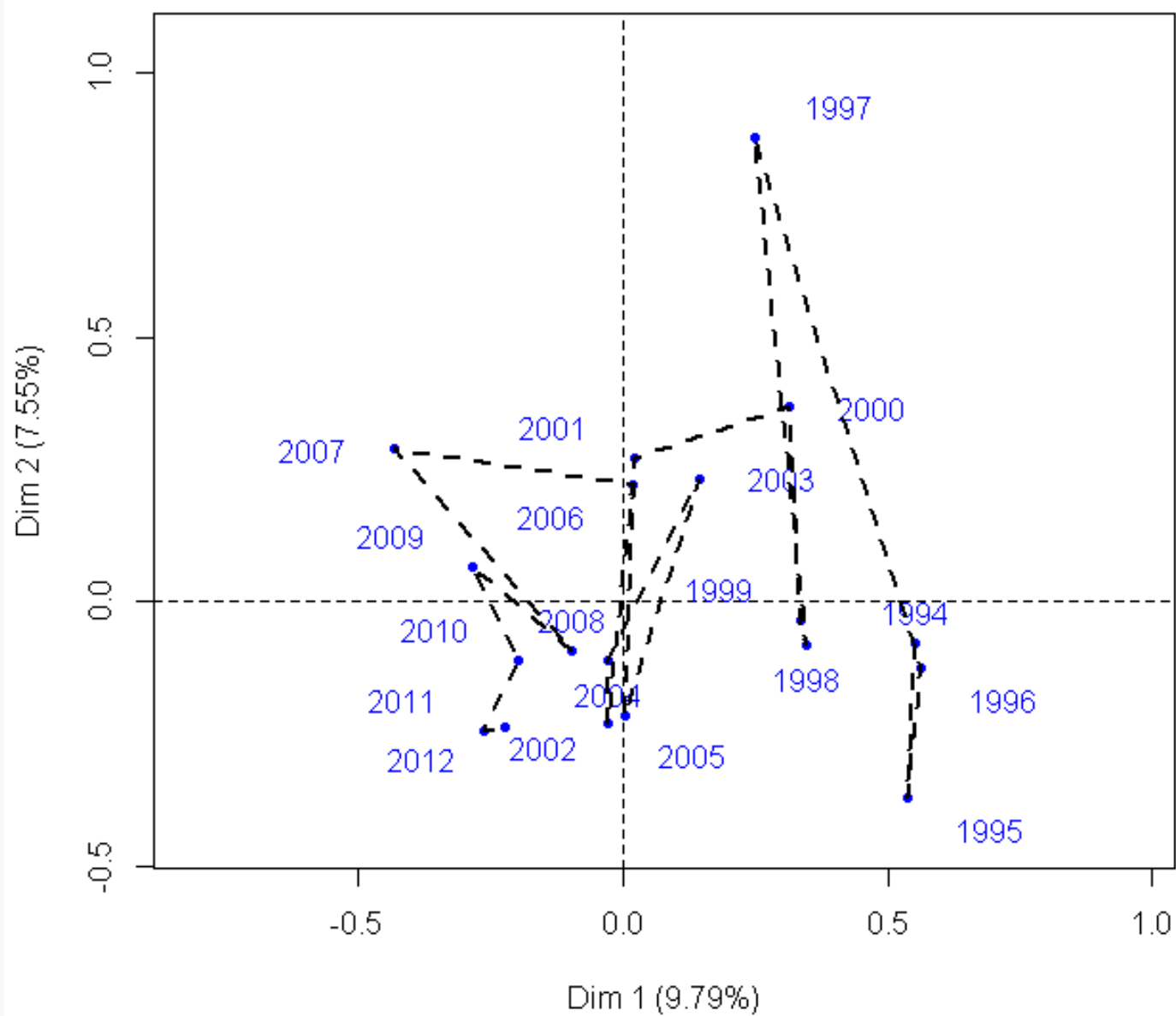
	dim 1	dim 2	dim 3	dim 4
Variance	0.07800	0.06019	0.05521	0.05094
% of var.	9.78603	7.55220	6.92749	6.59043
Cum %	9.78603	17.33823	24.26572	30.85615

6-1. Exemple : base Biblio

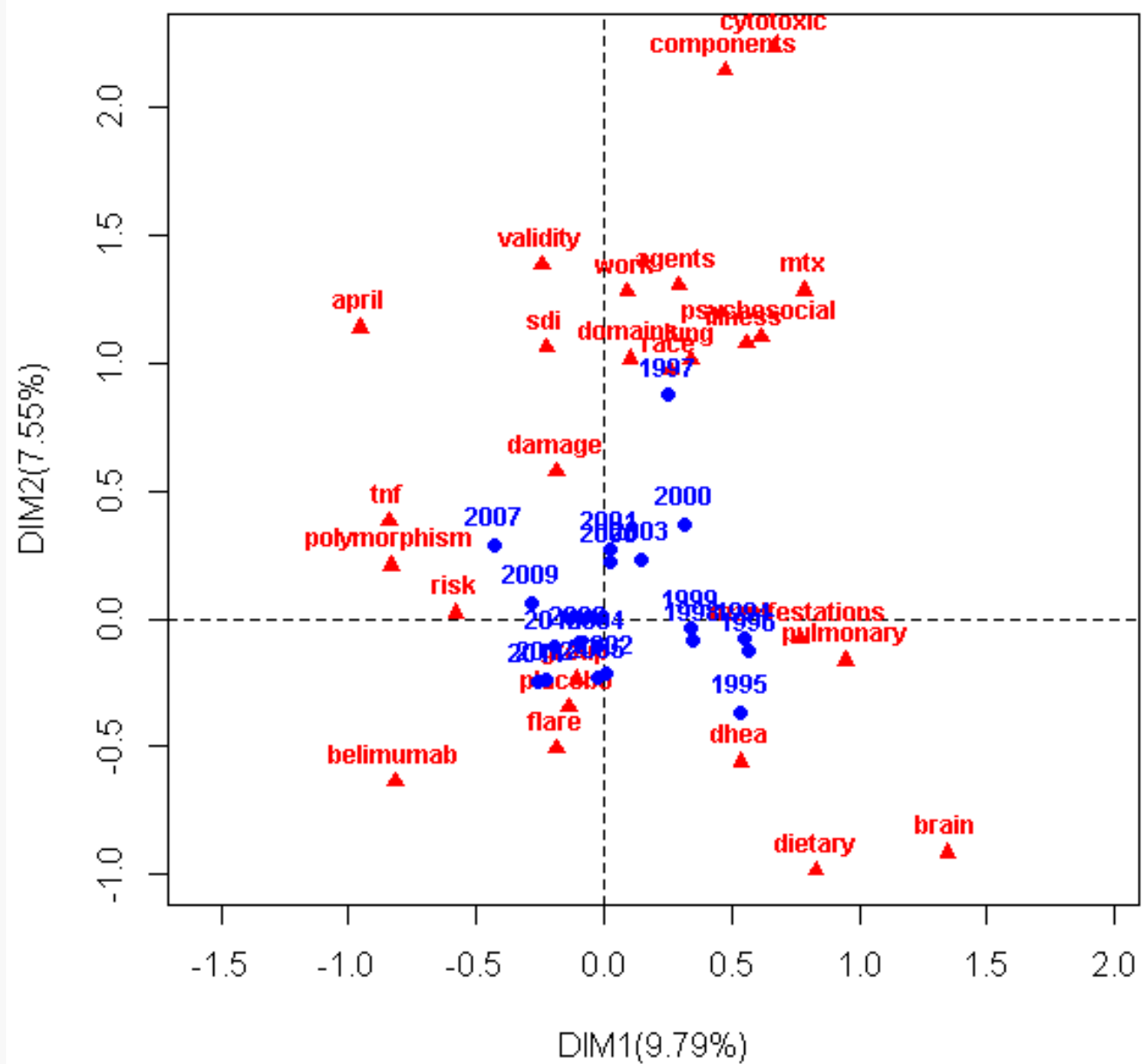
AC du tableau agrégé par année

- MacroBiblio fournit les mots caractéristiques de chaque année
- `$`2010`$Over_represented_word`
"events" "cognitive" "performance" "mri" "signals"
"mmf" "regimen" "cns" "independently" "newly"
"attributed" "atherosclerosis" "levels" "auc"
"creatinine" "variants" "european" "risk" "organ"
"multicenter« "diagnostic" "outcome" "older" "ldl"
- `$`2010`$Infra_represented_word`
"pregnancy" "antibodies" "cell" "dsdna"

CA documents

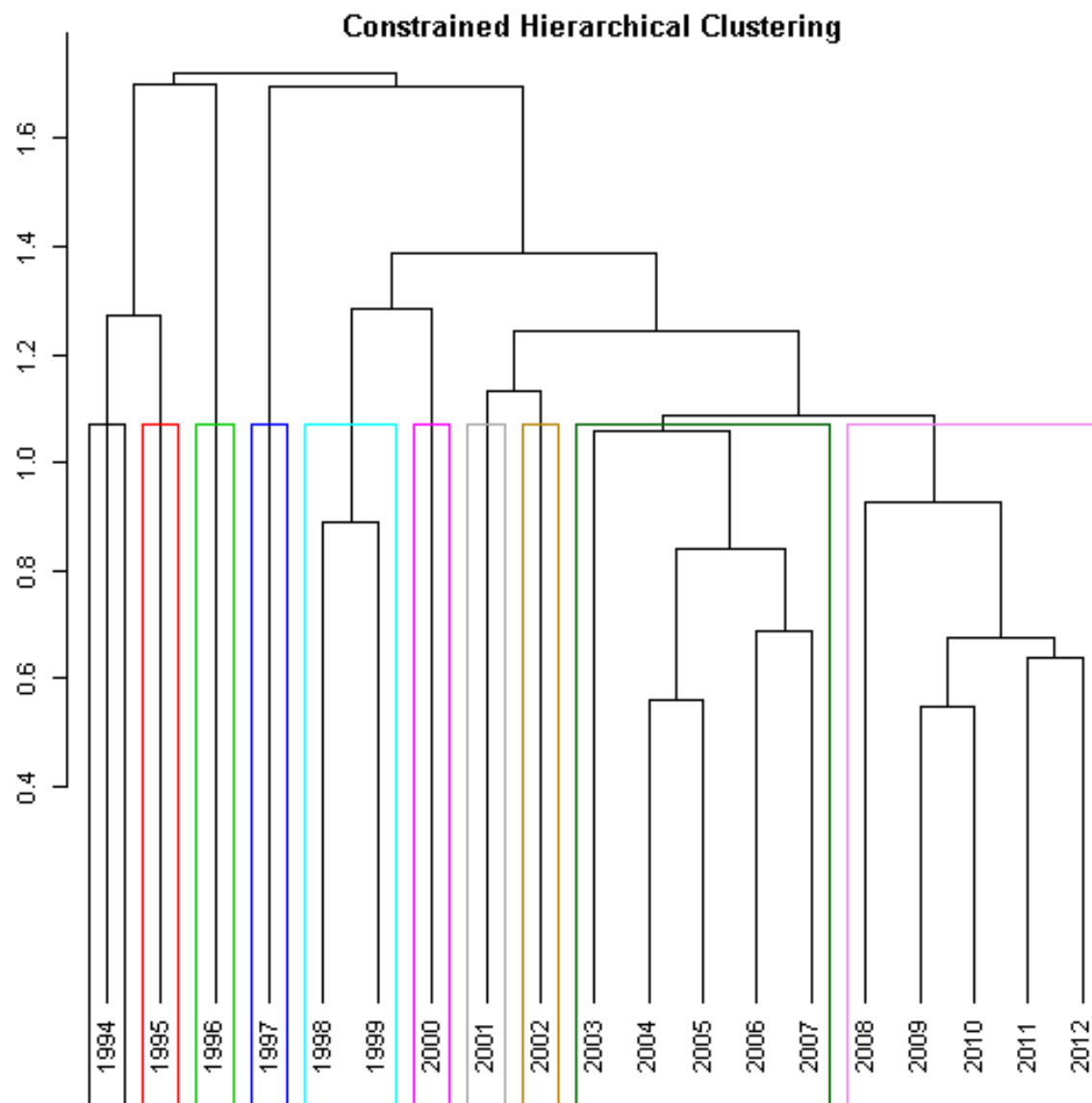


CA documents/words



Cl

Constrained Hierarchical Clustering



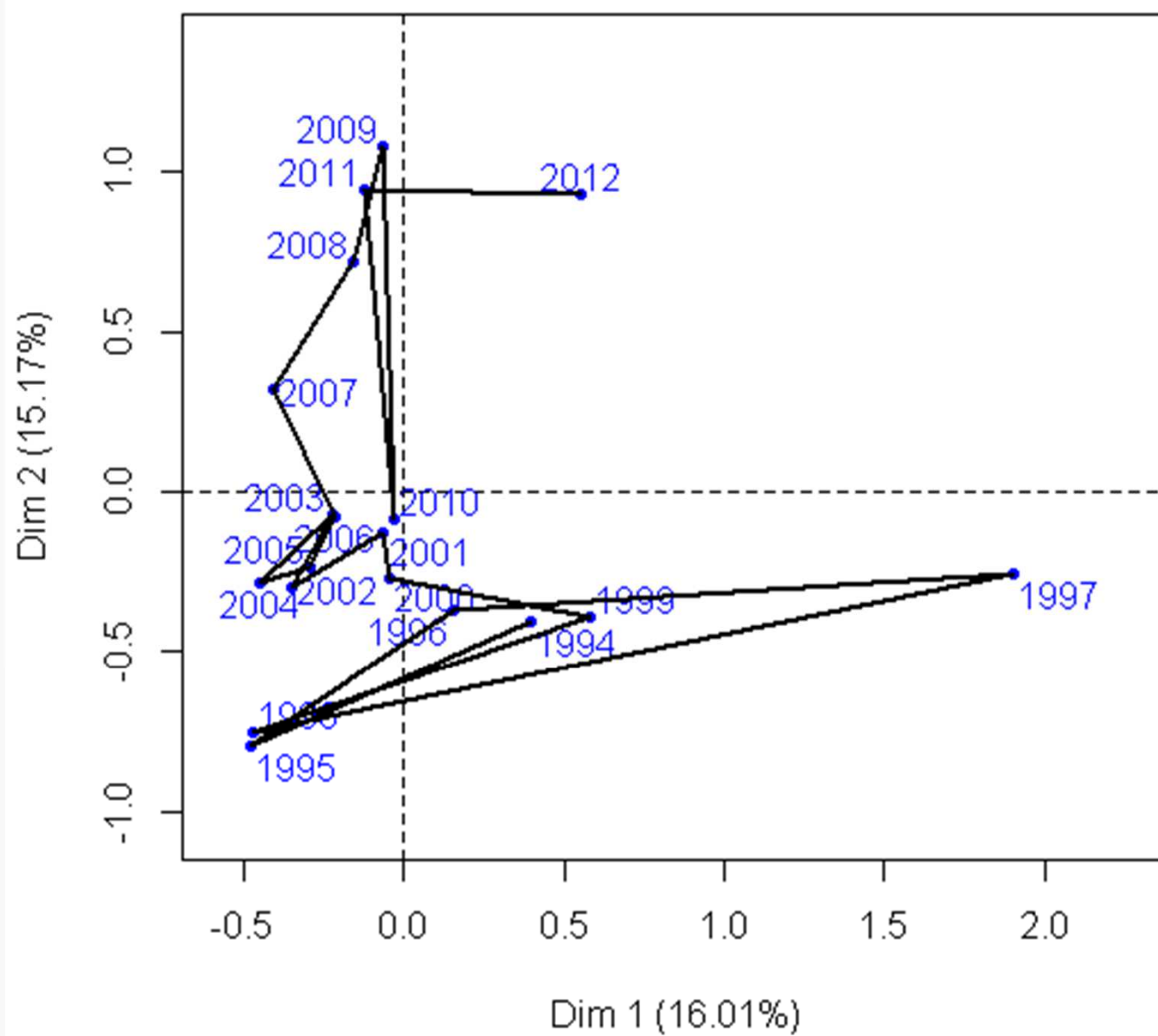
ilarity levels

AC sur noms de médicaments

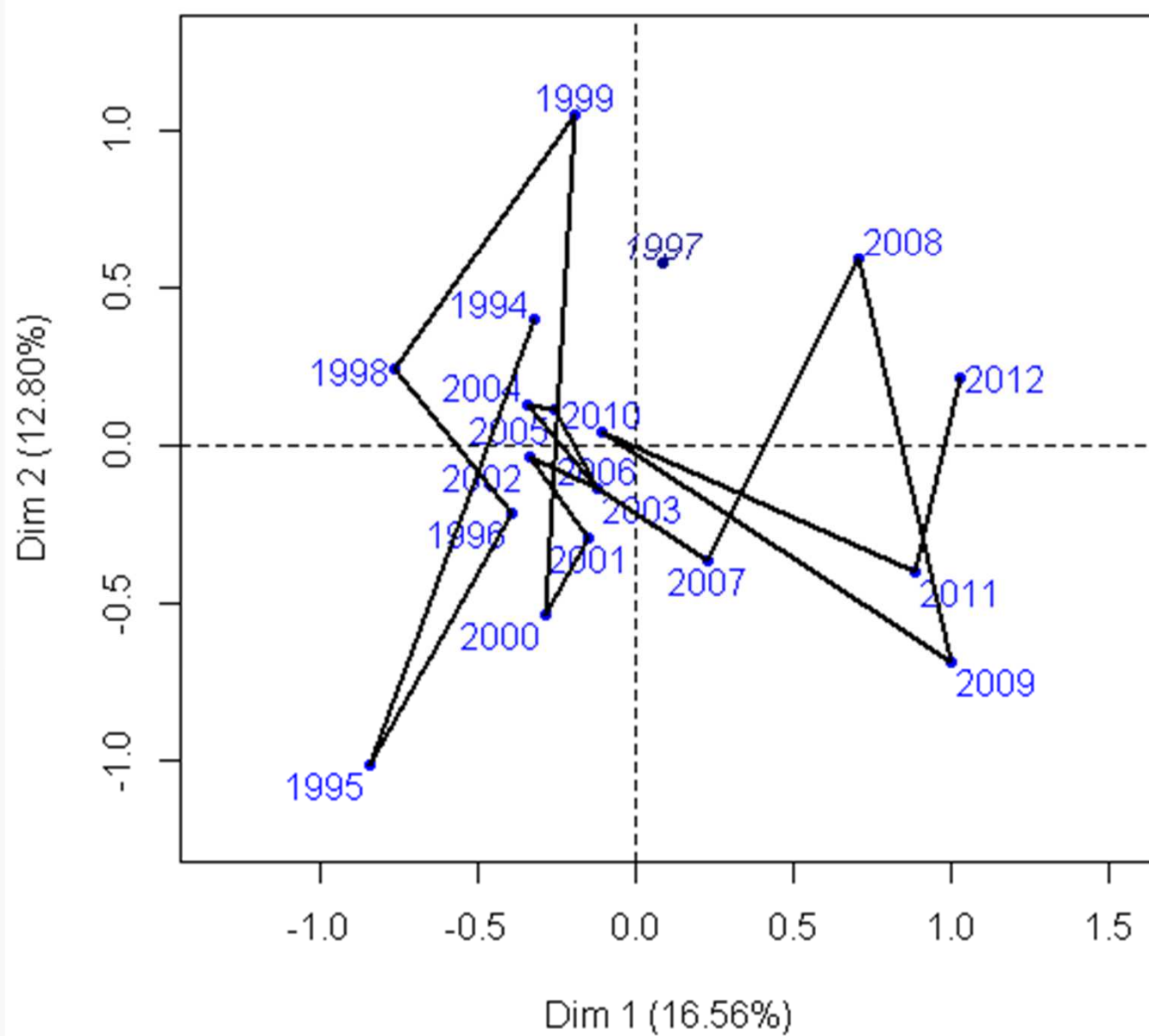
41 noms de médicaments et 19 années

- 1994 Corticosteroids
- 1995 bromocriptine prednisone anticardiolipin
- 1996 csa cyclophosphamide antimalarial anticoag antibody
- 1997 mtx cytotoxic methotrexate csa
- 1998 dheia hrt
- 1999 mtx ivig
- 2000 prednisone cyclophosphamide
- 2001 antibodies
- 2002 chloroquine prasterone
- 2003 cyclophosphamide
- 2004 dheia prasterone rituximab
- 2005 hrt antibody
- 2006 rituximab steroids
- 2007 rituximab infliximab
- 2008 abetimus methotrexate prasterone dheia
- 2009 infliximab rituximab
- 2010 csa 2011 belimumab antibody
- 2012 belimumab mtx plasma corticosteroids

CA factor map



CA factor map



6.1 Conclusion de l'analyse

- L'analyse des correspondances permet d'accéder au contenu des articles
- La classification hiérarchique permet de voir l'évolution de la recherche dans le temps
- L'étude sur les noms des médicaments fait apparaître deux périodes : avant 2007 et après 2007 : avant remixumab et bemimumab.