

L'analyse conceptuelle assistée par ordinateur et les humanités numériques

Louis Chartrand



Plan

I. Les humanités digitales

II. La LACTAO

III. Expérimentations

Partie I

Les humanités digitales

(vues par un philosophe des sciences)

Le télescope de Galilée



Extension de nos sens à travers de la lentille

La computation comme lentille

- Avancements en mathématiques : moteur des innovations scientifiques (Humphreys, 2004)
 - Production de théories
 - Ouverture du champs des investigations possibles
- Ainsi :

Télescope



La vue
(Précision, résolution, ...)

Informatique



La perception &
l'entendement
(Traitement de données)

Perception

- Passivement :
 - Emphase sur les propriétés intéressantes
 - Dans la vision, emphase des contrastes, reconnaissance de formes
 - Sélection d'attributs qui servent à nos tâches
 - Effets Sapir-Whorf
- Activement :
 - Recherche active des propriétés intéressantes

Augmentation quantitative

- Plus de données
 - Études de phénomènes très larges
 - Réduction du degré d'incertitude
- Plus de détails
 - Nouveaux éléments syntaxiques, stylistiques, sémiotiques, etc.
 - Liens entre niveaux
 - Relations entre niveaux textuels, ou avec d'autres données

Augmentation structurelle

- Façon alternative de structurer l'information
 - Biais différents sur l'information
 - Capacités de traitement de très nombreuses variables
 - Visualisations
- Produit un point de vue différent sur les données
 - L'hybride humain-machine produit un agent épistémique distinct

Bref...

- Les humanités digitales concernent :
 - Le renouvellement des outils (conceptuels, techniques) d'analyse
 - L'ouverture d'un champ empirique
 - L'augmentation de l'agence épistémologique
- Mais aussi :
 - Support matériel (édition, numérisation, lecture)
 - Création assistée par ordinateur
 - Etc.

Partie II

Lecture et analyse de texte assistée par ordinateur

(LACTAO)

Approches extrêmes

Traditionalisme

- L'analyse de texte requiert de l'interprétation
- La machine ne sait pas interpréter
- ∴ Donc, la machine ne peut aider l'analyse

Post-scientisme

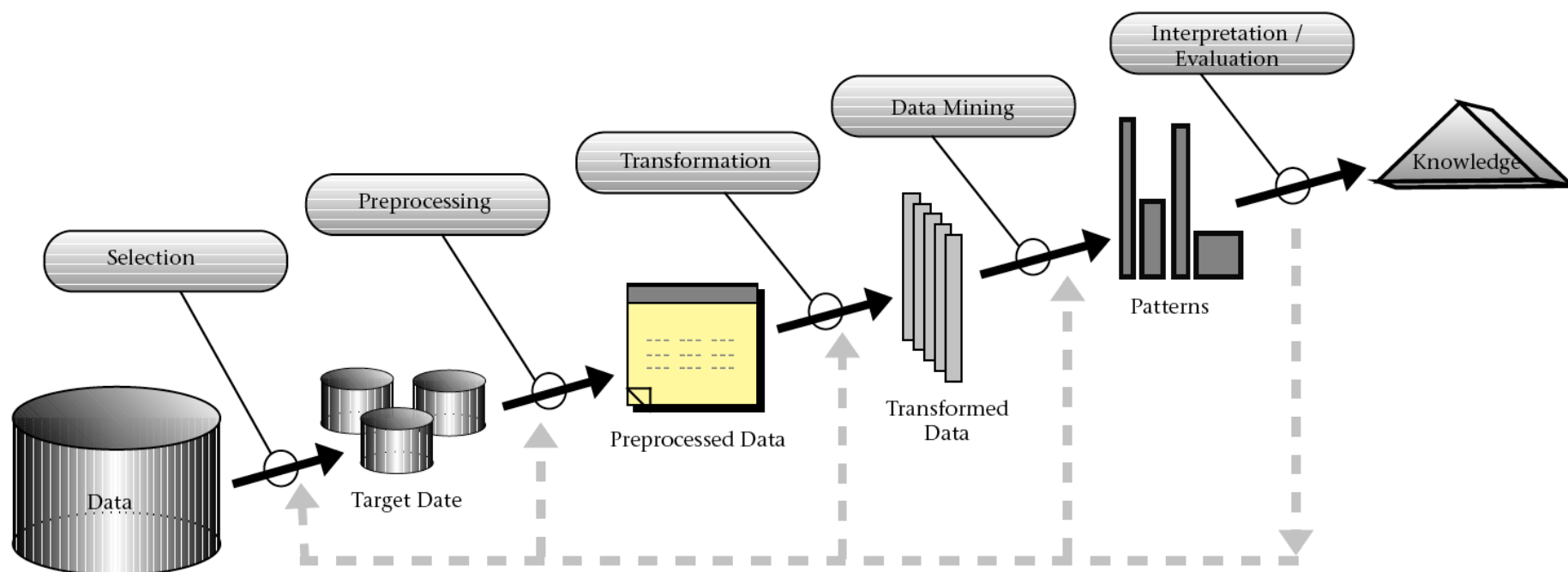
- Les données sont tellement massives qu'on peut détecter n'importe quelle régularité
- Il n'y a donc pas besoin d'interpréter les données
- ∴ Donc, on peut se passer des humains pour l'analyse

Approche d'assistance

Jean-Guy Meunier et Dominic Forest, « Classification and categorization in computer-assisted reading and text analysis », *Handbook of Categorization in Cognitive Science*, ed. H. Cohen et C. Lefebvre, 2005

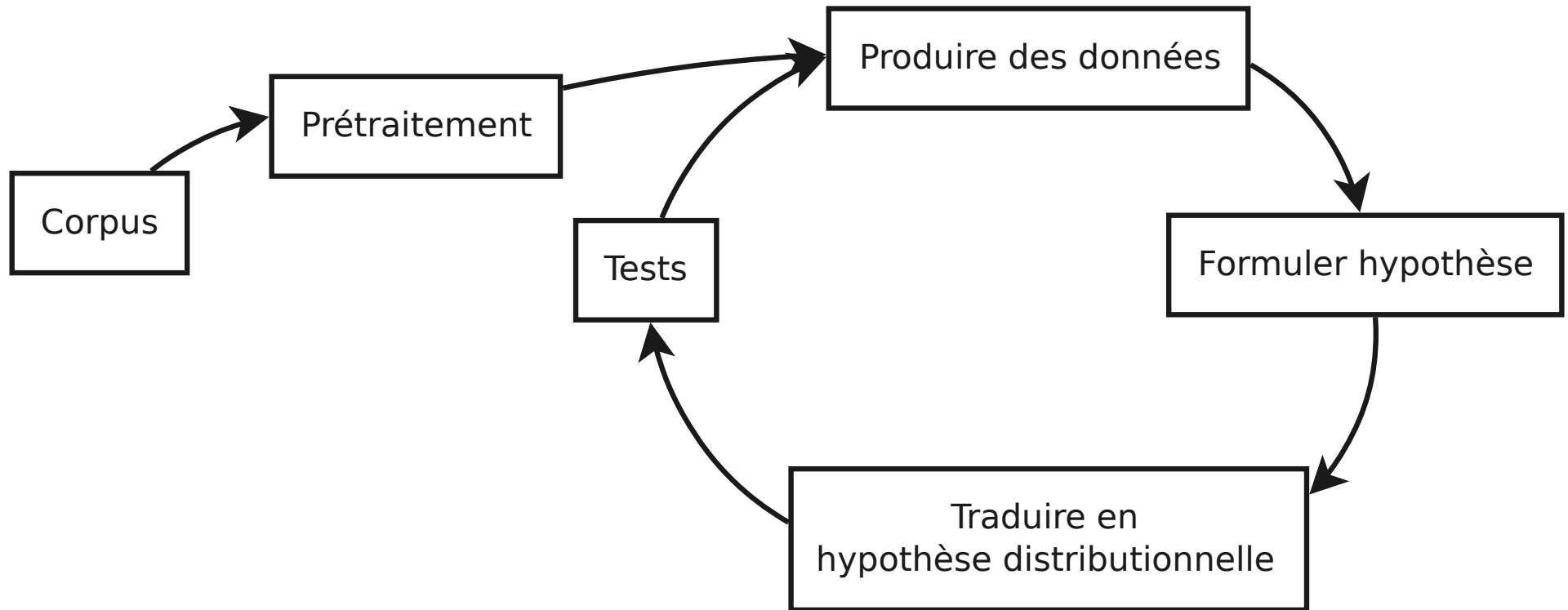
- Même une machine évoluée ne peut apprendre une langue qu'en voyant les mots
 - Cela évacuerait la sémantique (ancrage dans le monde) et la pragmatique (son usage)
- Les outils informatiques peuvent aider *certaines parties* du travail interprétatif
 - En particulier, la classification et la catégorisation

La chaîne de traitement



Fayyad et al. « From data mining to knowledge discovery in databases », *AI Magazine*, 1996

Le processus de recherche

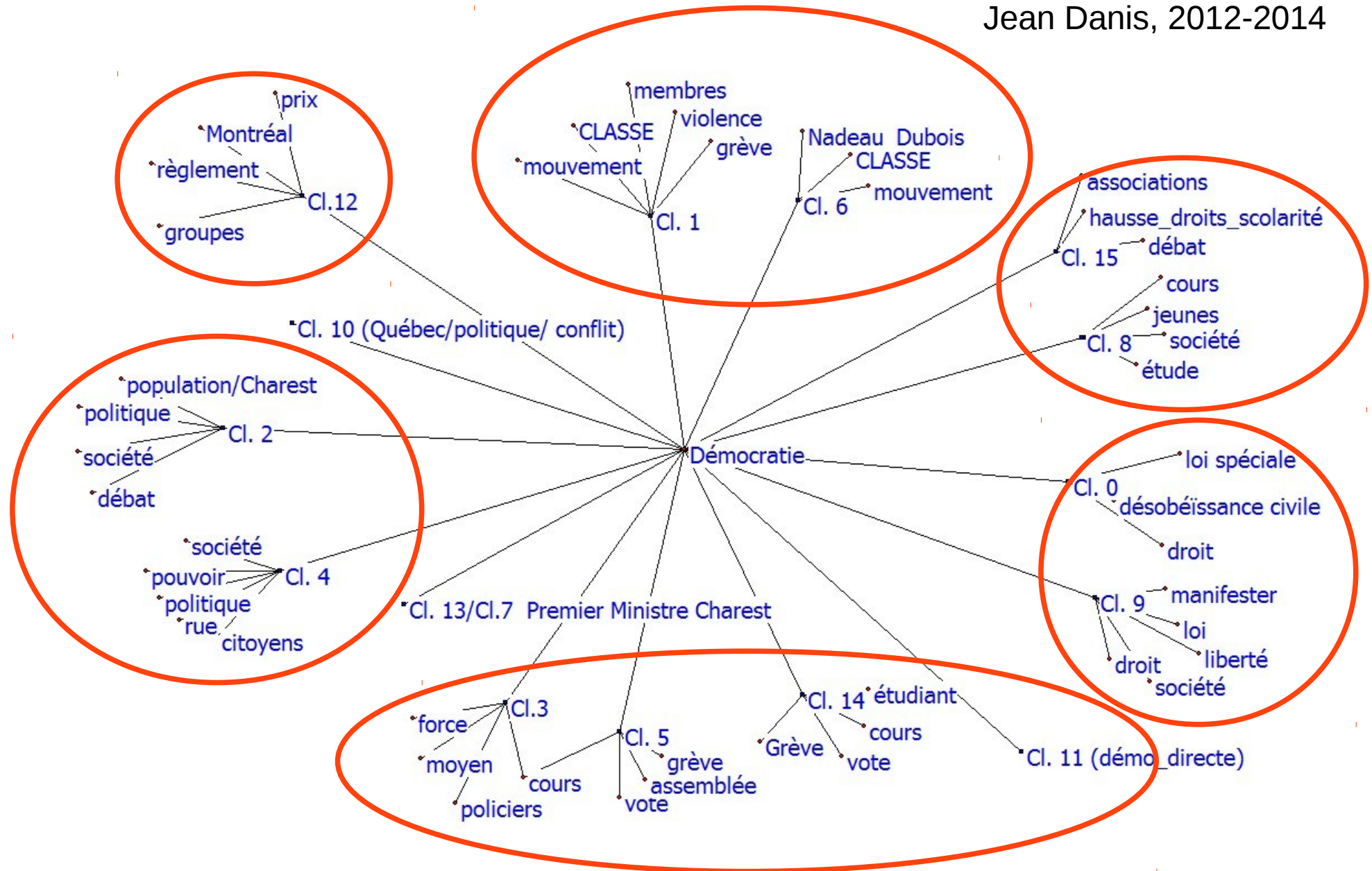


Partie III

Expérimentations

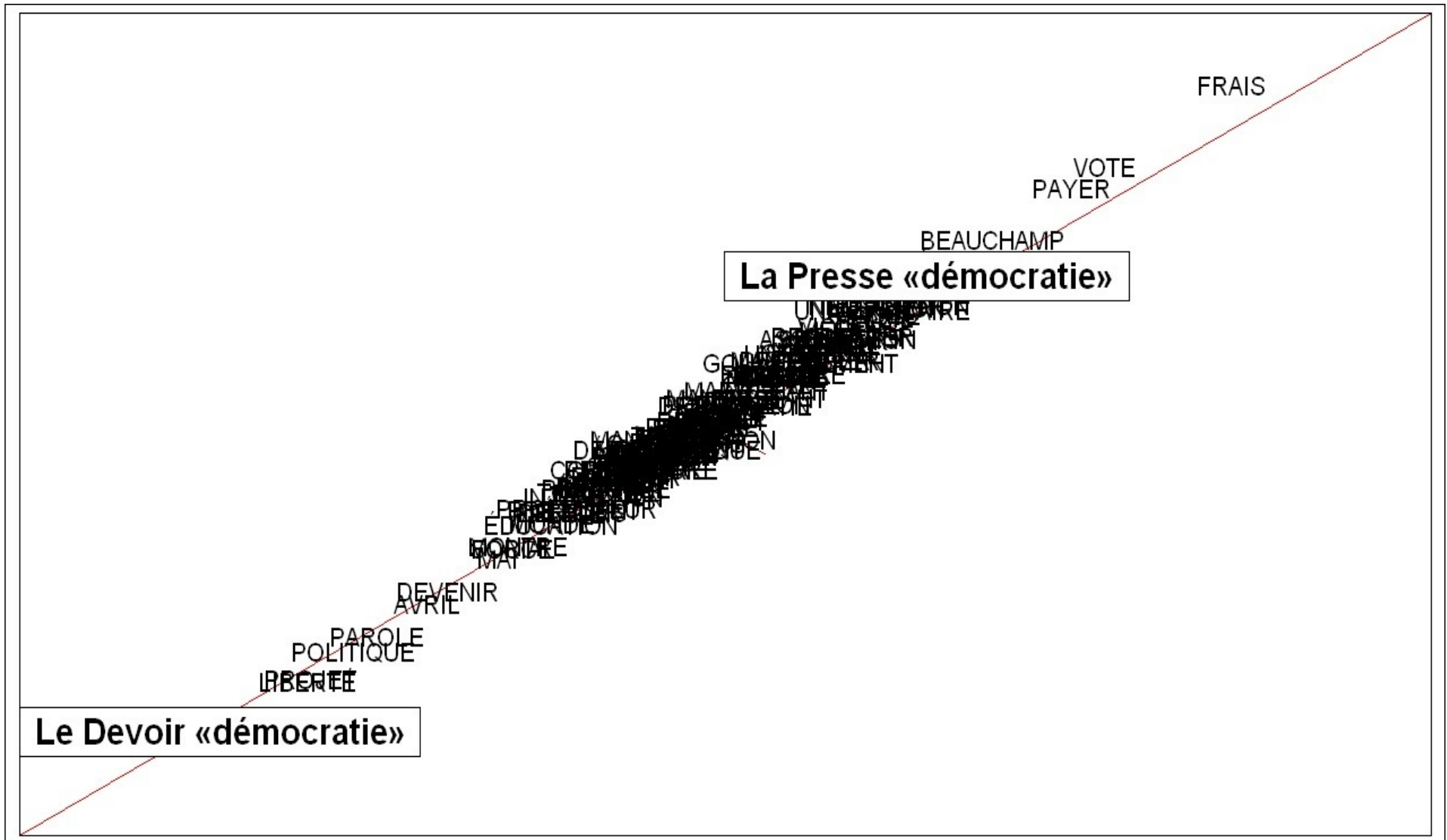
« Démocratie » dans le printemps étudiant

Jean Danis, 2012-2014

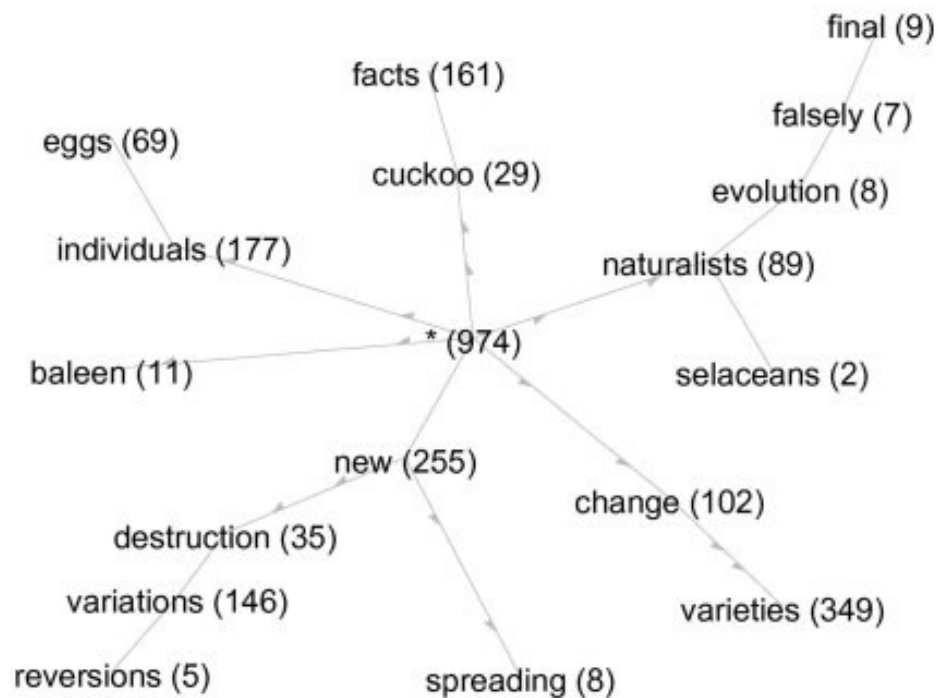


« Démocratie » dans le printemps étudiant

Jean Danis, 2012-2014



Le concept d'évolution chez Darwin



- Le mot « evolution » est excessivement rare chez Darwin
- Solution :
 - Prendre les mots fortement associés au contexte où se retrouve « évolution »
 - Faire de même avec les mots trouvés
 - Analyser l'arbre obtenu
- Les « pistes » correspondent à différentes façons de parler d'évolution

Maxime Sainte-Marie, Jean-Guy Meunier, Nicolas Payette et Jean-François Chartier, « The concept of evolution in the Origin of Species: a computer-assisted analysis », *Literary and Linguistic Computing*, 2011

Annotations sur l'œuvre de Magritte

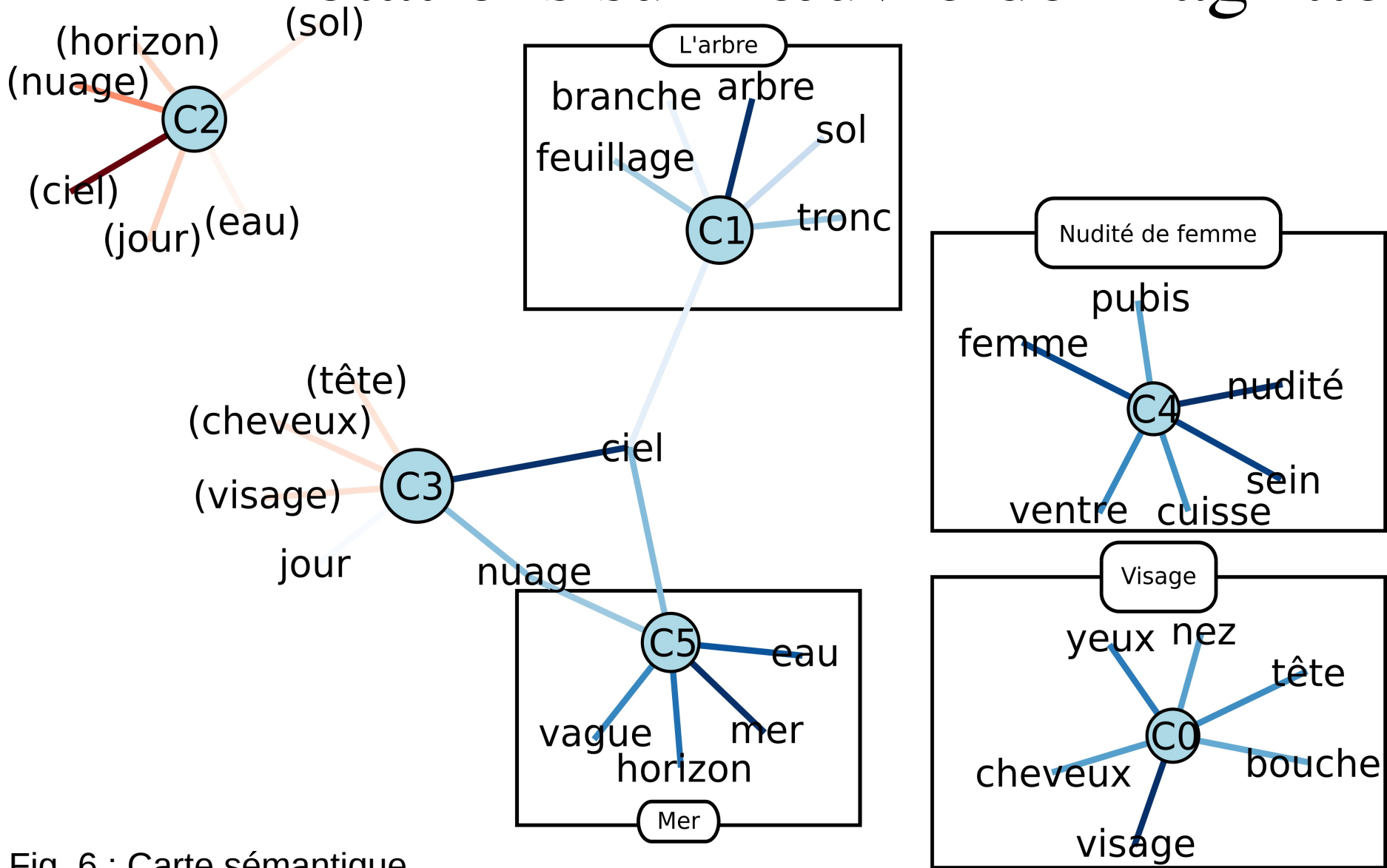


Fig. 6 : Carte sémantique représentée sous la forme d'un graphe en étoile

Louis Chartrand et Jean-Guy Meunier, « Peindre Magritte avec des mots », *Cahier de l'ISC*, à paraître

Distributions temporelles

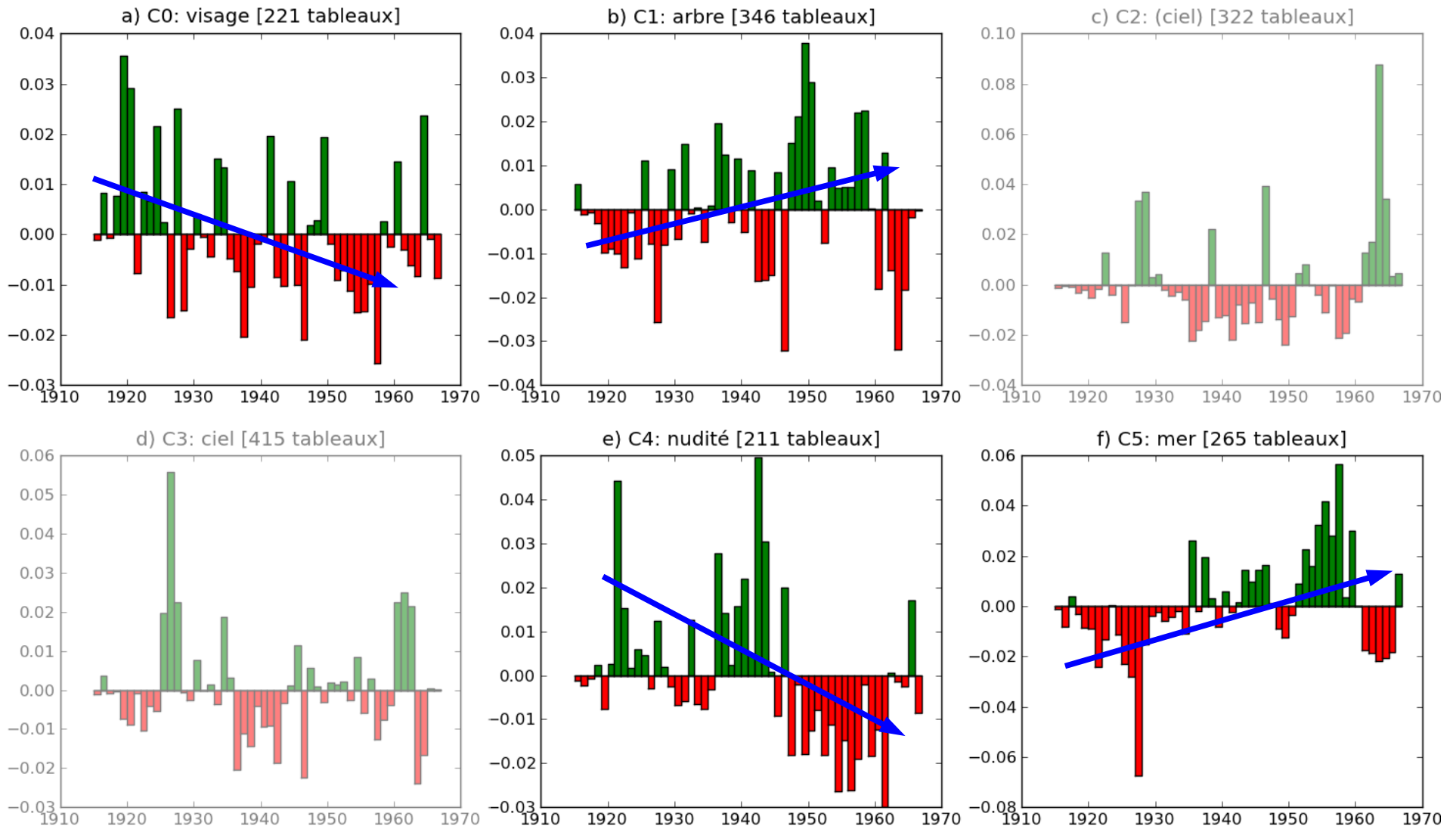


Fig. 7 : Distribution temporelle des six classes. La valeur en ordonnées est la différence entre la proportion d'œuvres appartenant à la classe en question dans une année donnée et la moyenne de la même statistique pour les autres classes.

Le librisme

Lila Roussel, 2014

- Étude sur des articles sur le logiciel libre et le librisme, en majorité de la *Free Software Foundation*
- Question : Promotion de liberté positive ou négative?

Attributs associés aux concepts de liberté :

Liberté positive	Liberté négative
Communauté	Individu
Justice distributive	Libre marché
Collaboration	Compétition
Bien commun	Propriété privée
Capacité à agir	Liberté de choix
Concept d'exercice	Concept d'opportunité
Libre des contraintes internes	Libre des contraintes externes

Le librisme

Lila Roussel, 2014

Cooccurrents de
« Freedom »

Total No. of Collocate Types: 435		Total No. of Collocate Tokens: 9336			
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	1307	76	76	11.51315	freedom
2	82	44	38	9.04912	free
3	116	66	50	7.94060	software
4	44	34	10	6.41807	value
5	41	12	29	6.32206	one
6	41	14	27	5.93196	community
7	36	29	7	5.87700	give
8	32	10	22	5.43464	price
9	33	6	27	5.37186	help
10	30	4	26	5.26993	run
11	49	18	31	5.14630	program
12	33	25	8	5.10504	computer
13	27	25	2	5.03383	wish
14	27	9	18	4.94643	make
15	28	1	27	4.86543	distribute
16	25	3	22	4.82158	redistribute
17	24	24	0	4.72681	respect
18	24	0	24	4.69536	study
19	28	12	16	4.65853	use
20	44	32	12	4.60370	people
21	22	18	4	4.52097	take
22	26	5	21	4.50577	change
23	19	6	13	4.25099	zero
24	21	1	20	4.08882	publish
25	22	6	16	3.87085	source
26	16	16	0	3.74656	live
27	29	17	12	3.67889	user
28	16	11	5	3.66344	goal
29	15	0	15	3.55470	share
30	12	11	1	3.35875	defend
31	12	0	12	3.34704	cooperate
32	12	5	7	3.32597	neighbor
33	12	7	5	3.29320	essential
34	11	6	5	3.28239	like
35	12	5	7	3.27681	anu

Le librisme

Lila Roussel, 2014

Cooccurents de
« freedom and »
→

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	18	0	18	3,53155	community
2	5	0	5	1,98944	cooperation
3	3	1	2	1,69927	free
4	4	2	2	1,63100	give
5	3	1	2	1,22168	instead
6	3	3	0	1,14677	live
7	3	3	0	0,90797	value
8	3	3	0	0,67854	talk
9	4	0	4	0,32530	use
10	4	4	0	-2	your
11	4	1	3	-2	you
12	3	2	1	-2	won
13	7	1	6	-2	we
14	4	2	2	-2	way
15	3	2	1	-2	want
16	3	3	0	-2	values

Concordance de « not freedom » ↓

good. Alternative possible masters is not freedom. And competition could lead them to do a "better" job.
. Controlling other people's lives is not freedom at all, not for either of the people involved. 35. Can
harms society. A choice of masters is not freedom. Discussions of rights and rules for software have oft
between a discrete set of options is not freedom. Freedom is something much deeper and much broader than
of masters, and a choice of masters is not freedom. Freedom is having no master. 12. Copyleft forbidding
es will teach users that the point is not freedom. Nonfree software in GNU/Linux distros already works ap
to A, because to subjugate others is not freedom. That's power. Now, there may be people who would like
ur master. But a choice of masters is not freedom, with free software you don't have to choose a master.
r free": it implies the issue is price, not freedom. One way to avoid the confusion is to say "release
rms imply that the issue is about price, not freedom. Some common terms such as "piracy" embody opinio
m to dictate to everyone!" But that's not freedom, that's power. So I'm making the distinction between t
tended this to happen. Their goal was not freedom, just "success", defined as "having many users." Th