

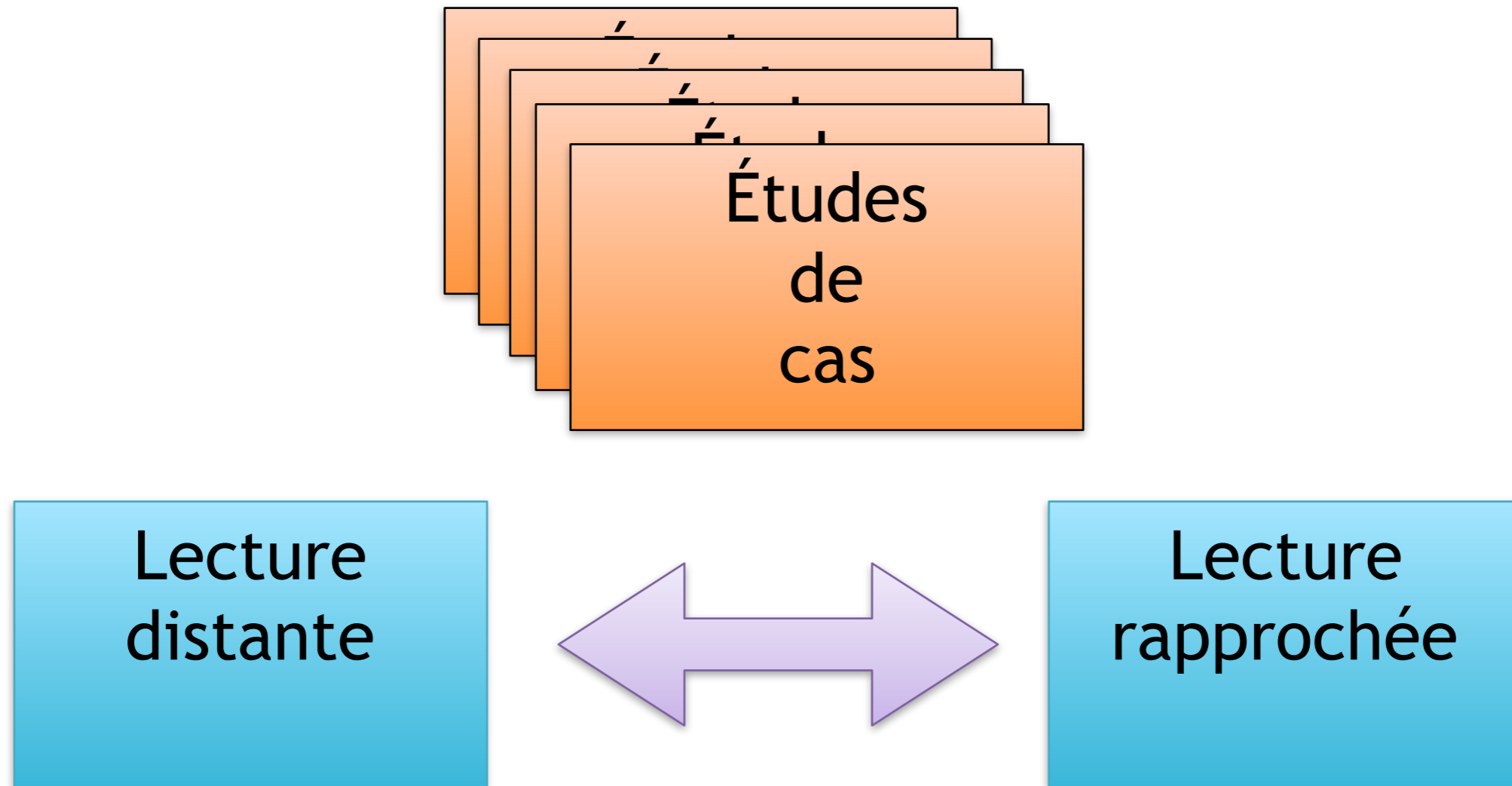
Forer l'imprimé du XVIIIe siècle français

Un journal de guerre en 70 000 volumes

François Dominic Laramée

Chercheur postdoctoral en histoire numérique de la médecine
Chaire Jason A. Hannah de l'histoire de la médecine, Université d'Ottawa
30 janvier 2019

CADRE DE RECHERCHE - MÉTHODOLOGIE



A) Tendances dans la production imprimée

- **Numérique > Lecture traditionnelle**

B) Vastes corpus apparentés

- **Numérique \sim Lecture traditionnelle**

C) Petits corpus fortement interreliés

- **Numérique < Lecture traditionnelle**

CADRE DE RECHERCHE - MODÈLES THÉORIQUES

RICHARD H. THALER
WINNER OF THE NOBEL PRIZE IN ECONOMICS
and CASS R. SUNSTEIN
WINNER OF THE HOLBERG PRIZE



Nudge

NEW YORK TIMES Bestseller

Improving Decisions About
Health, Wealth, and Happiness

"One of the few books . . . that fundamentally changes the way I think
about the world." —Steven D. Levitt, coauthor of FREAKONOMICS

MICHEL FOUCAULT

**L'ordre
du discours**

Gérard Genette
Introduction
à l'architexte

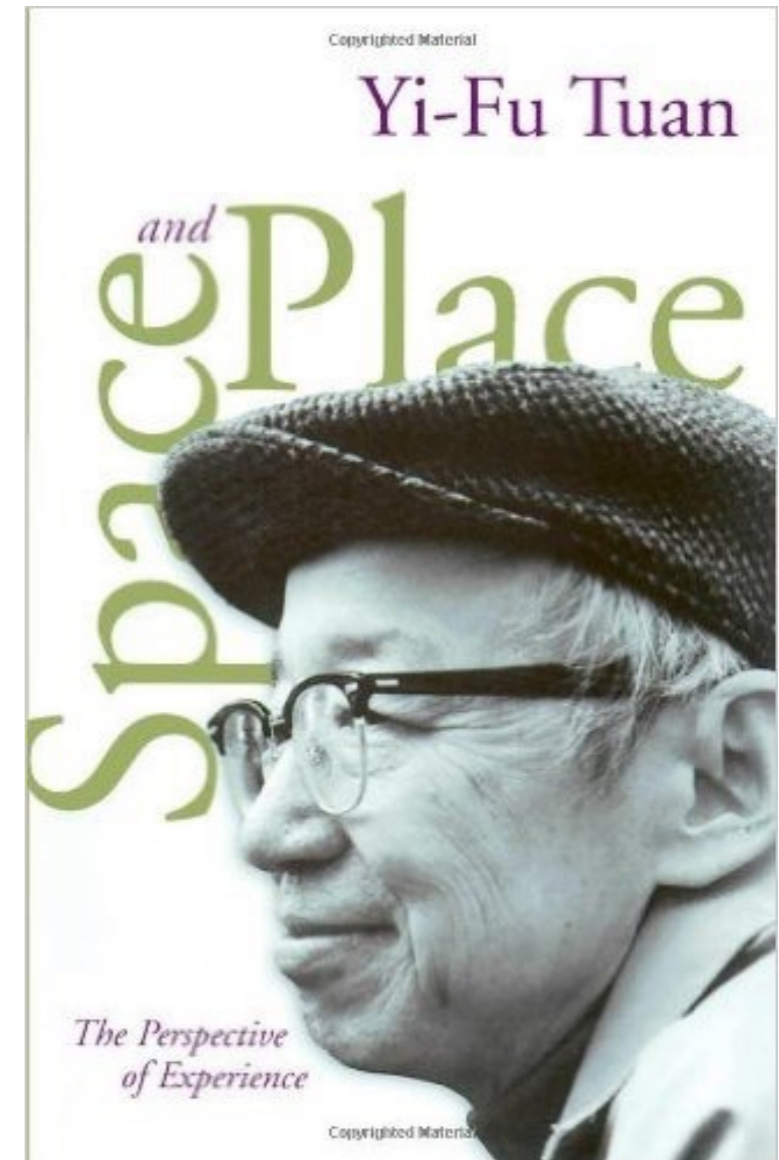
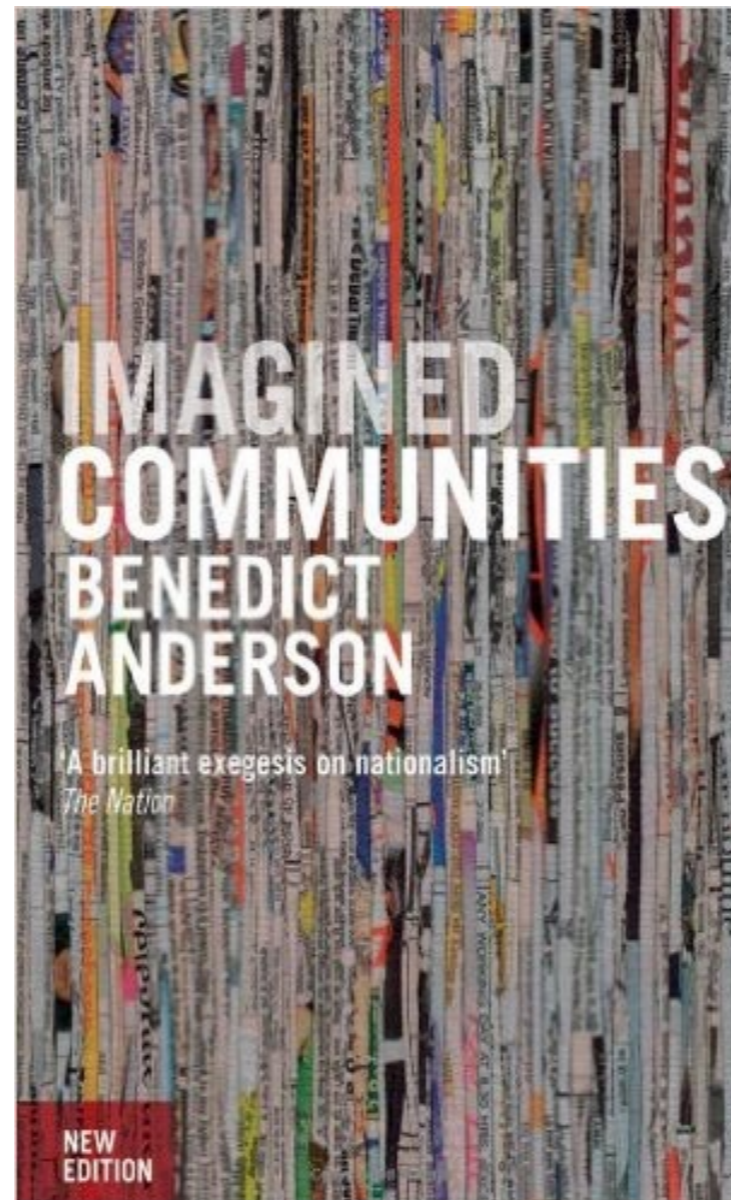
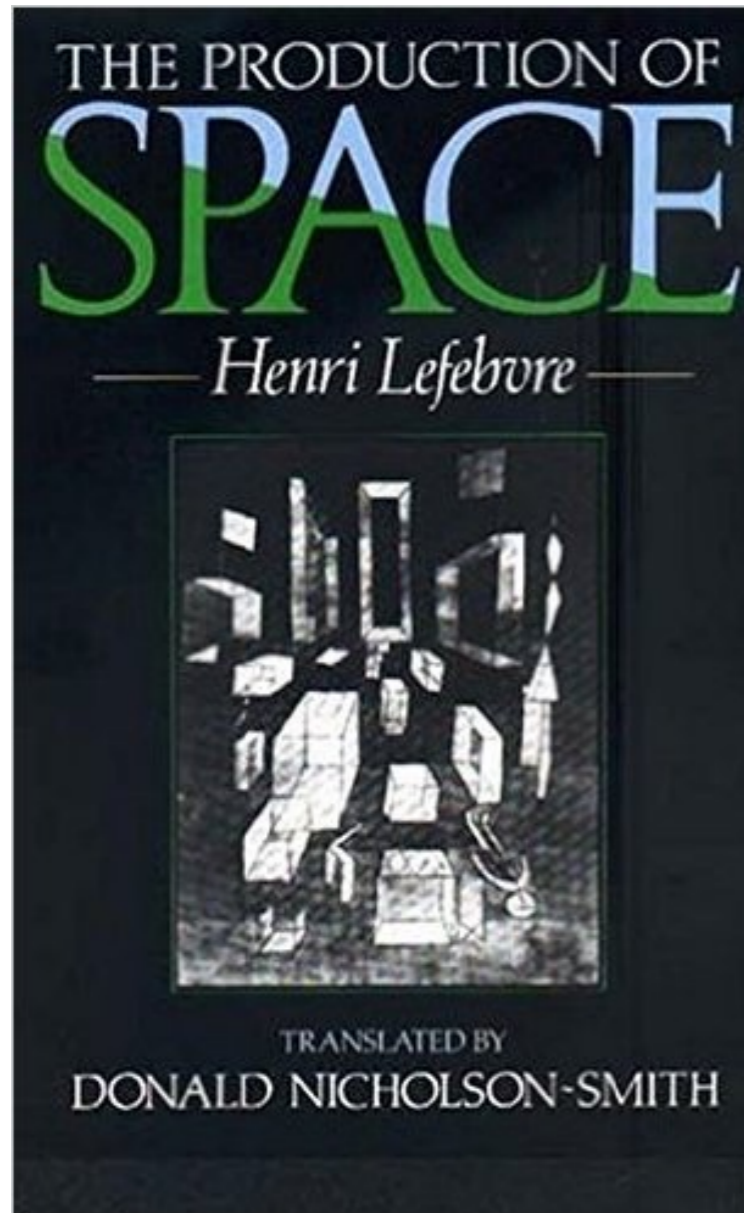
THINKING,
FAST AND SLOW



DANIEL
KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS

CADRE DE RECHERCHE - GÉOGRAPHIE IMAGINÉE



« La stratégie est un système de pis-aller. »

— Helmuth von Moltke l'Ancien (1800-1891)



LES TROUPES DE L'ENNEMI

18. LOUYSIANE, la (page 9:707) [Géographie] Jaucourt

LOUYSIANE, la LOUYSIANE, la, (Géog.) grande contrée de l'Amérique **longe**, ou les Inc

en septentrionale & en méridionale par le golfe de
Louïsiane, la Virginie, le Canada, Terre- neuve,

Don Joseph Zugnol, Premier Médecin du Roi, aura la direction de cet établissement.

Le 1^r du mois dernier, on ressentit à Maroc un affreux tremblement de terre, à la même heure qu'en Espagne. La plupart des maisons & des édifices publics de cette Ville ont été totalement renversés, & une grande multitude d'habitans a été enseveli sous les ruines. Le nombre des personnes, qui ne peuvent pas encore se fixer. A huit lieues de c

L-e 1^r du mois demih, On reRèntic à Maroc un affreux '- ViMe même beure qu'en Efpa < > ne, La plâpart des imifons & des-édificies publics de ctteVille ont été totalement renver * une grande mul,": titude d'habitans aéreenfevelie fous les ruines. Le nom- PfS encore sebre des personnes, qui ne peut pas encore [e ftrer].

MODÈLES VS RÉALITÉ



Rappelez-vous que tous les modèles sont faux; la véritable question consiste à savoir à quel point ils doivent être faux pour ne plus être utiles.

— George E. P. Box, statisticien
(*photo par David Eddy, wikipedia*)

Faiblesses

- Fiabilité?
- Représentativité?

Solution / précaution

- Approche *exploratoire* plutôt que *prédictive*
- Extraire (ou construire) des métadonnées
- Itérer, itérer, itérer

MÉTHODES NUMÉRIQUES EXPLORATOIRES vs INTERPRÉTATION



« L'expression la plus excitante dans le monde scientifique n'est pas : 'Eureka!', mais plutôt : 'Tiens, c'est curieux...' »

— Isaac Asimov

(Portrait par Rowena Morrill, Wikimedia)

CORPUS DE TRÈS GRANDE TAILLE : HATHI TRUST EXTRACTED FEATURES



HATHI
TRUST
Digital Library

Search the HathiTrust Digital Library

Full-text Catalog

Full view only

[Advanced full-text search](#)

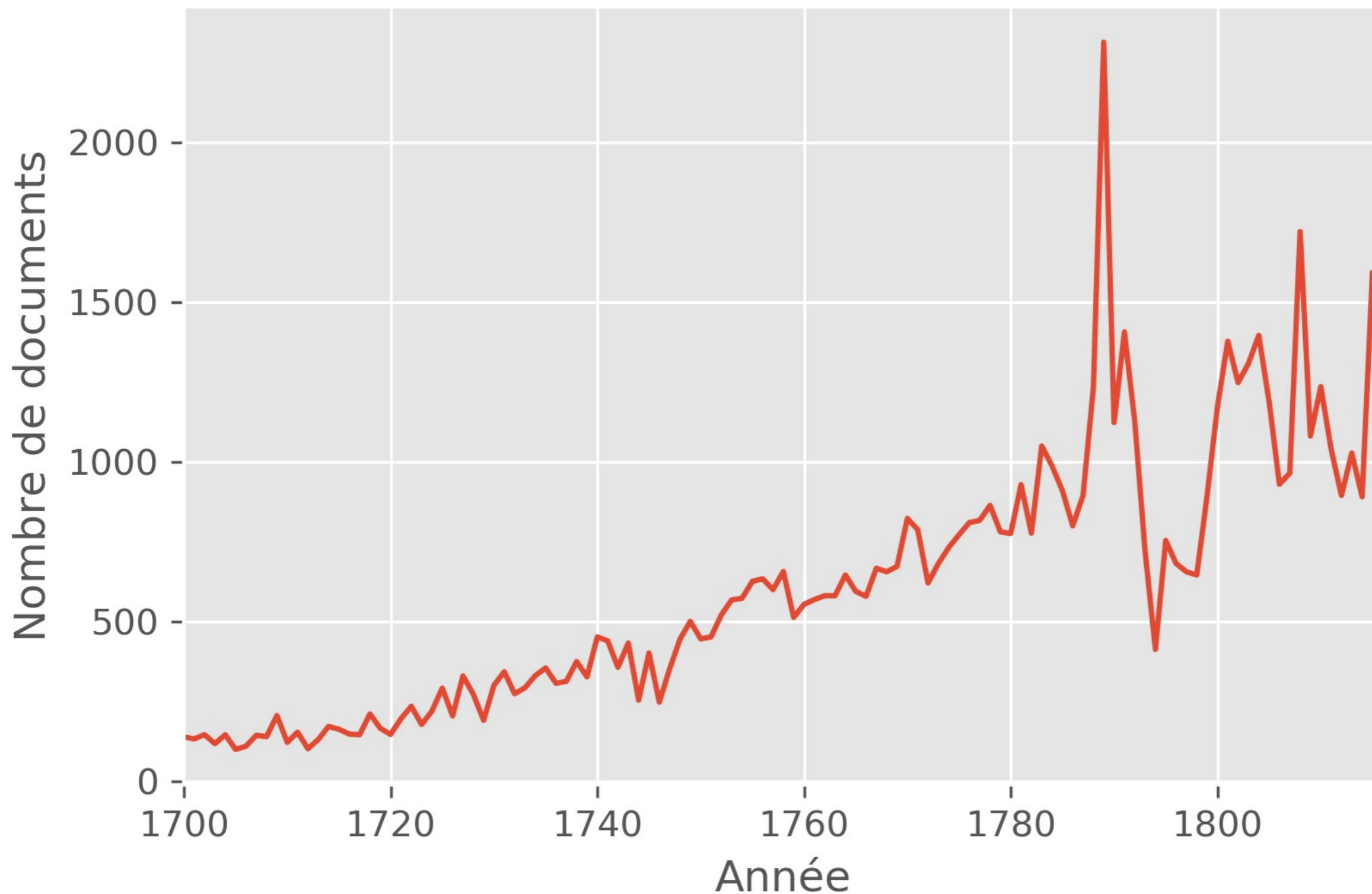
[Advanced catalog search](#)

[Search tips](#)

[*Should I search catalog or full-text?*](#)

LE JEU DE DONNÉES HATHI TRUST EXTRACTED FEATURES

Hathi Trust Extracted Features - Documents en français

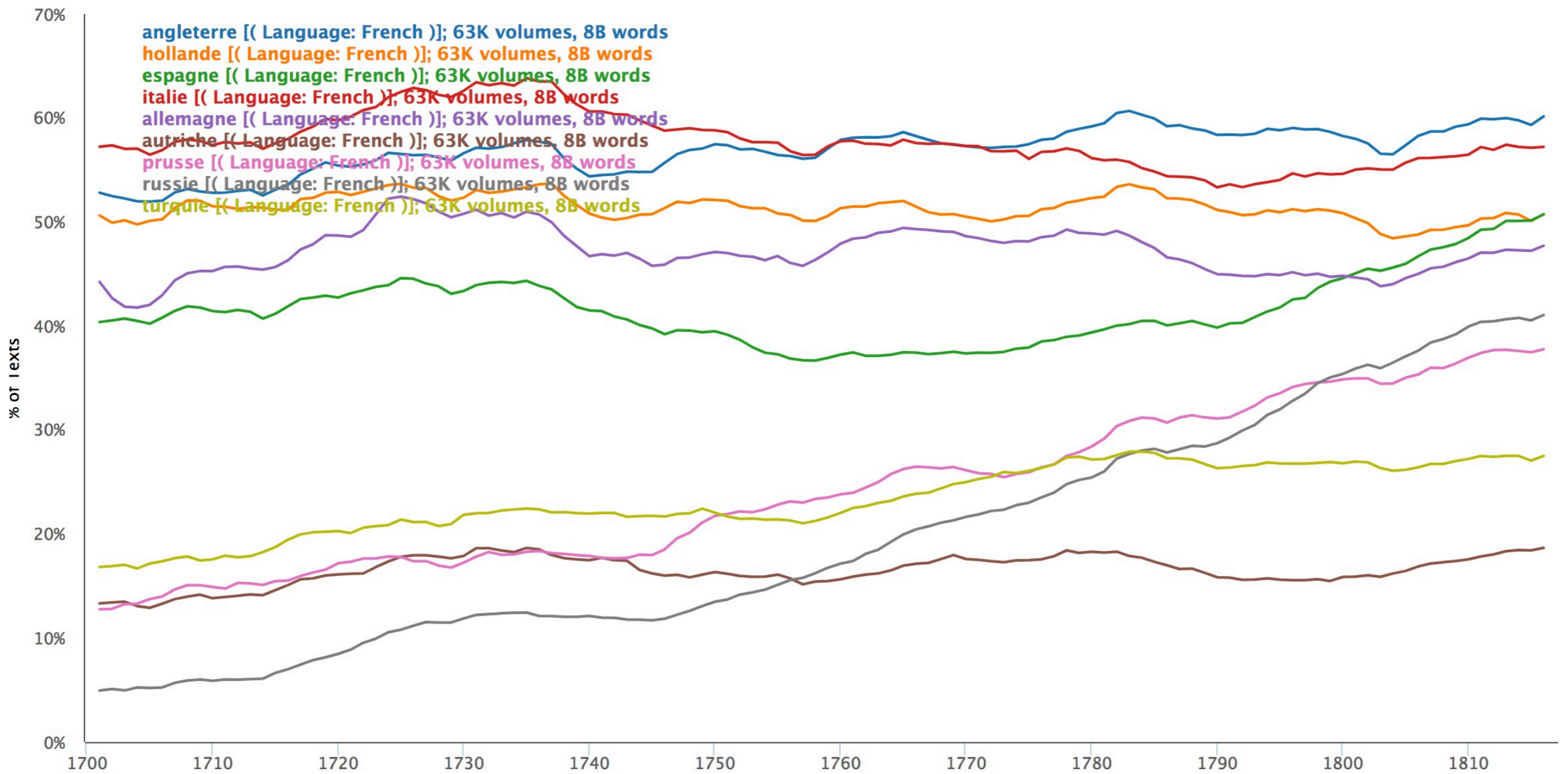


LES «SACS DE MOTS» DU HTEF

Source: *The Programming Historian*

				count
page	section	token	pos	
27	body	those	DT	1
		within	IN	1
28	body	a	DT	3
		be	VB	1
		deserted	VBN	1
		faintly	RB	1
		important	JJ	1

HATHI TRUST BOOKWORM



Quel contenu est pertinent?

- Règle heuristique: 3 mentions d'un pays dans la même page
- 380 ensembles de pages: un par grande puissance par année (Angleterre, Autriche, Russie, Espagne et Prusse; 1740-1815)
- Modèles calculés avec le module *Gensim* de Python

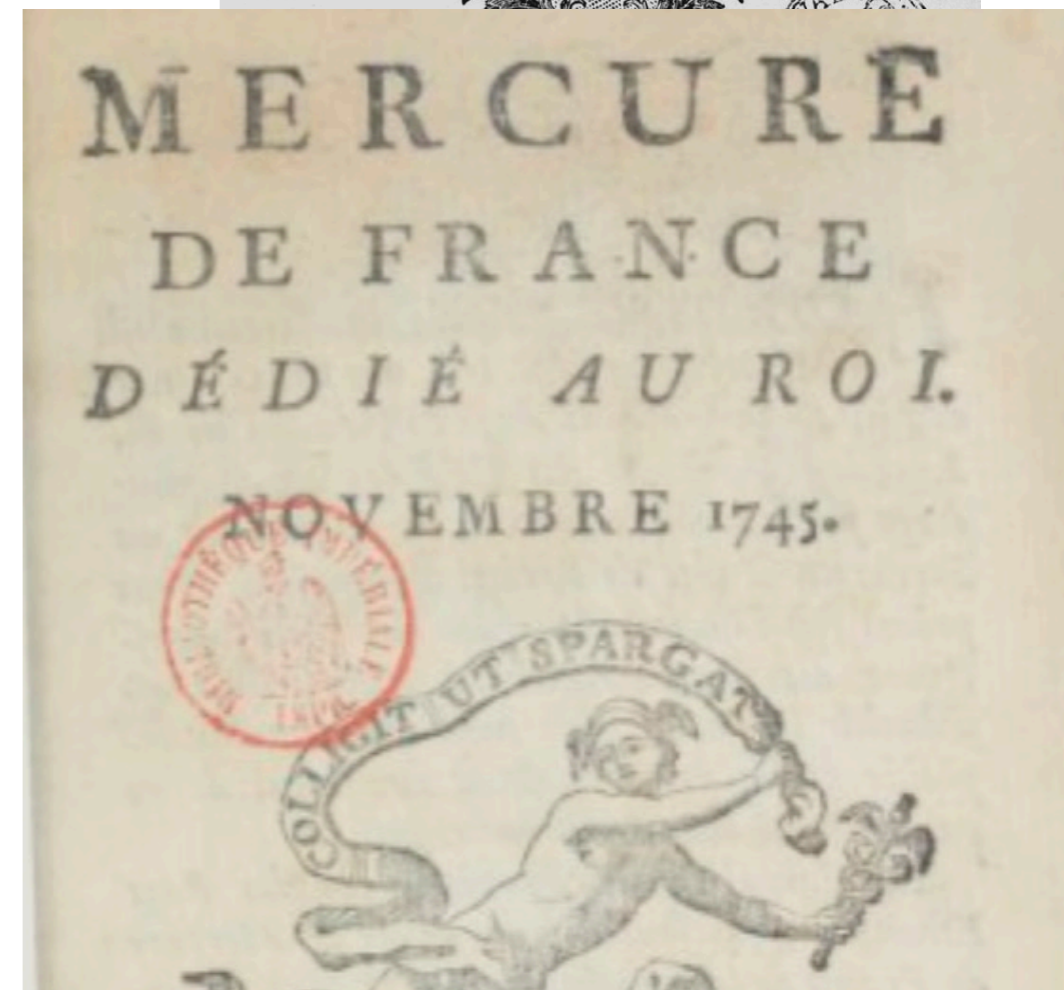
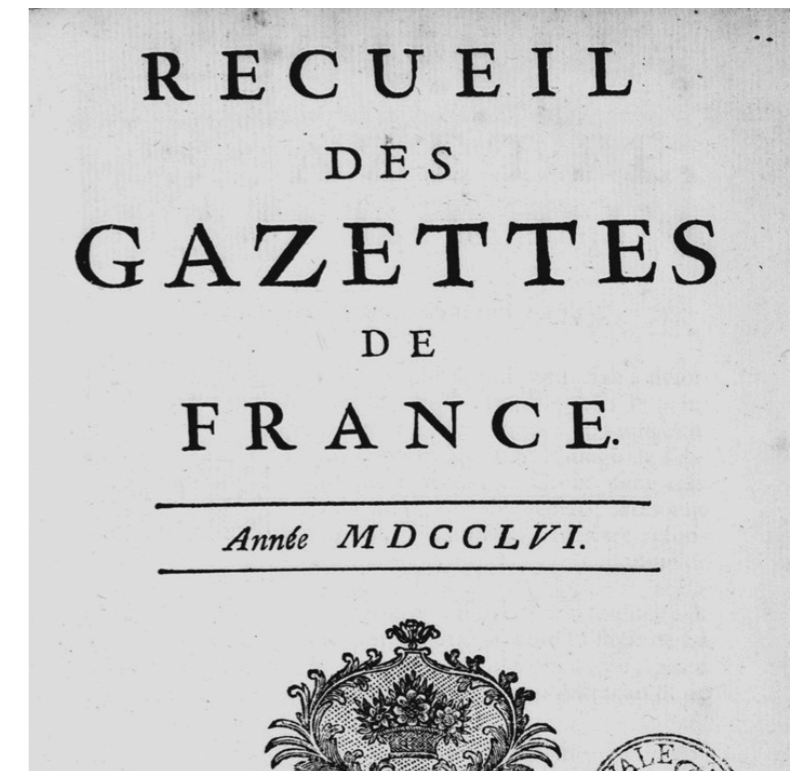
MODÈLE THÉMATIQUE — ANGLETERRE (extraits)

Sujet	Mots les plus importants	Sujet dominant (années)
12	roi france général point français ville pays grand partie	1805, 1808-09, 1811, 1814
6	roi france français anglais général guerre ville mort paris	1807, 1810, 1812, 1815
10	roi france duc prince point françois guerre pape troupes ville	1744, 1755-58, 1769, 1771
13	roi ville duc france prince compte l'an nom mort	1742-43, 1745-47, 1749
7	france roi guerre traité point général paix français hommes	1778, 1790, 1792-97, 1802, 1813

CORPUS DE LA PRESSE PÉRIODIQUE

Cinq périodiques (océrés):

- Gazette de France (nouvelles, 1740-1792)
- Mercure de France (généraliste, 1740-1758)
- Journal des Savants (culturel, 1751-1791)
- Journal des débats (nouvelles, 1800-1805)
- Journal de l'Empire (nouvelles, 1805-1815)



L'AUTRICHE DANS LE *JOURNAL DE L'EMPIRE*

Rang	Mot(s)	Rang	Mot(s)
1	France et synonymes	11	corps
2	l'empereur	12	général
3	Russie et synonymes	13	Allemagne et synonymes
4	troupes	14	Prusse et synonymes
5	Vienne	15	empire
6	comte	16	cour
7	Autriche et synonymes	17	roi
8	armée	18	maison
9	guerre	19	ici
10	prince	20	paix

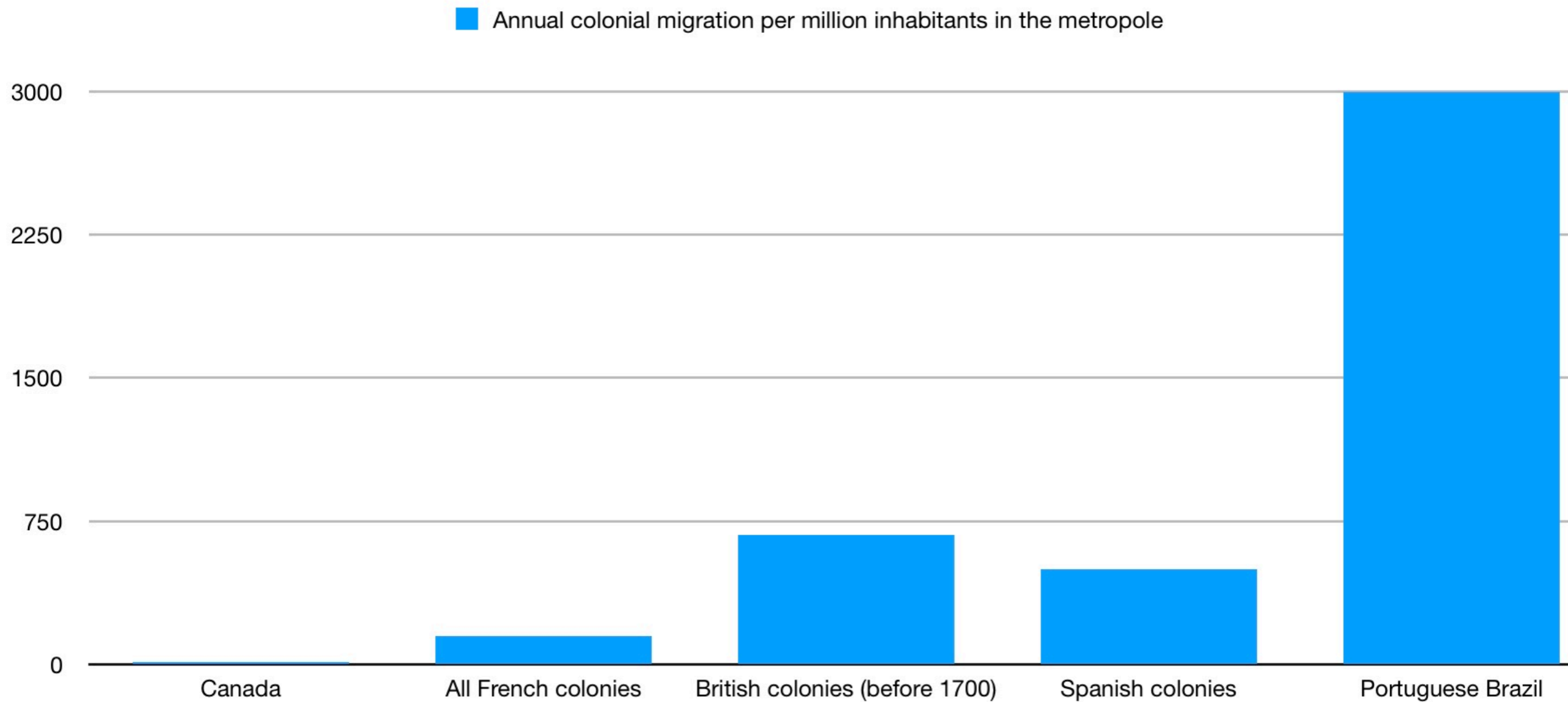
SIMILITUDE DES DISCOURS DANS LES JOURNAUX APPARENTÉS

Pays	Distance angulaire (Mercure / Gazette)	Distance angulaire (Journal des débats / de l'Empire)
Autriche	0,16	0,16
Angleterre	0,19	0,09
Espagne	0,16	0,15
France	0,25	0,26
Prusse	0,08	0,10
Russie	0,17	0,12

DISCOURS STÉRÉOTYPÉ DANS LES PÉRIODIQUES

Périodique	Moyenne	Médiane	Écart-type
Mercure de France	0,57 (0,46)	0,57 (0,46)	0,17 (0,12)
Gazette	0,47 (0,40)	0,49 (0,39)	0,15 (0,11)
Journal des débats	0,44 (0,32)	0,42 (0,33)	0,15 (0,08)
Journal de l'Empire	0,35 (0,25)	0,35 (0,25)	0,11 (0,07)
Journal des Savants	0,59 (0,59)	0,59 (0,64)	0,19 (0,20)

MIGRATIONS COLONIALES SOUS L'ANCIEN RÉGIME

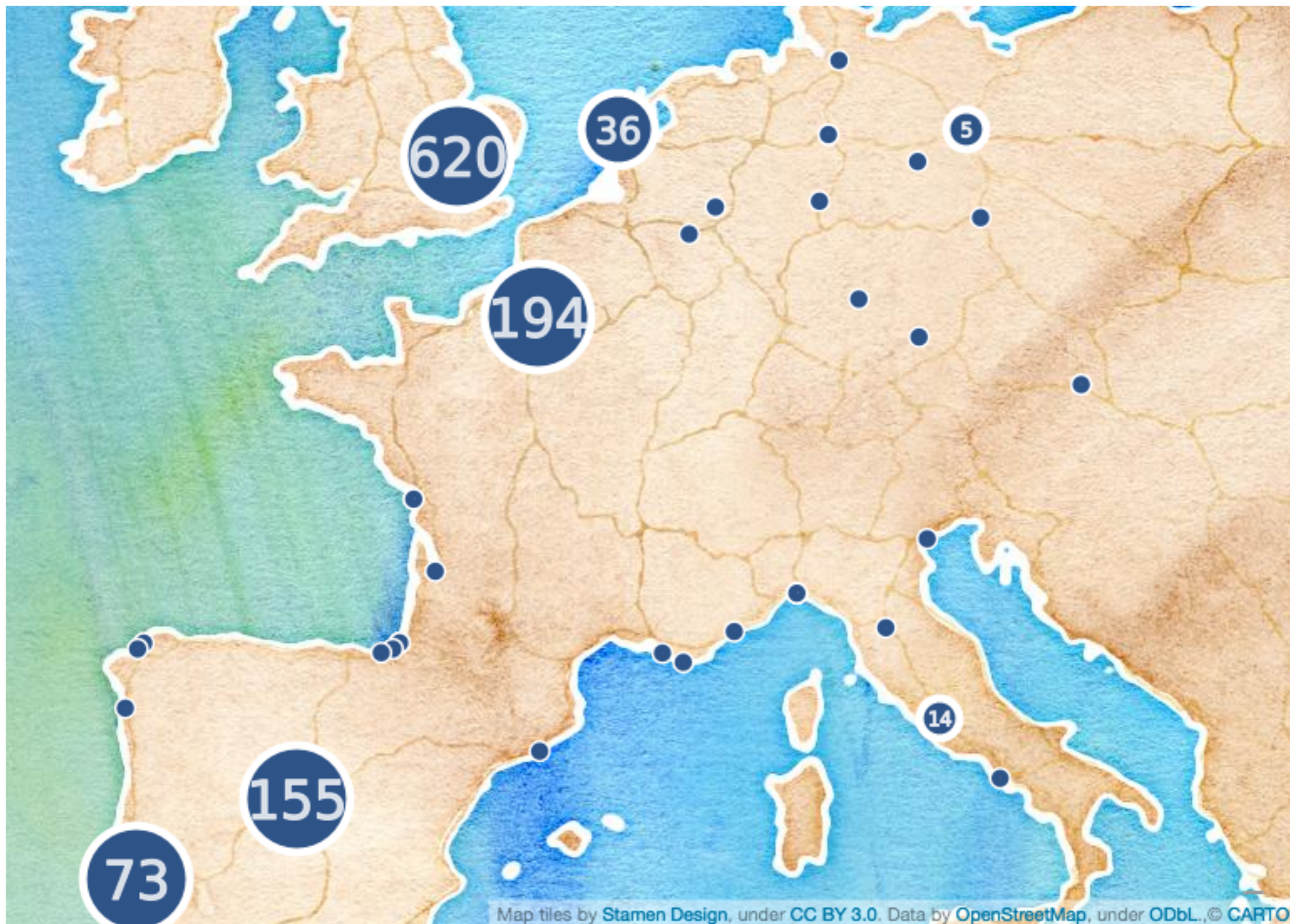


Données tirées de Landry (2006), « Les Français passés au Canada avant 1760 »

MÉTADONNÉES D'ARTICLES DE LA GAZETTE DE FRANCE

id	"ville"	"annee_pub"	"amerique_yn"	"louisiane_yn"	"colonie_yn"	"canada_yn"
am1740_1	Londres	1740	1	0	0	0
am1740_10	Madrid	1740	1	0	0	0
am1740_11	Londres	1740	1	0	0	0
am1740_12	Londres	1740	1	0	0	0
am1740_13	Dresde	1740	1	0	0	0
am1740_14	Madrid	1740	1	0	0	0
am1740_15	Londres	1740	1	0	1	0
am1740_16	Londres	1740	1	0	0	0
am1740_17	Londres	1740	1	0	0	0
am1740_18	Madrid	1740	1	0	0	0
am1740_19	Londres	1740	0	0	1	0
am1740_2	Madrid	1740	1	0	1	0

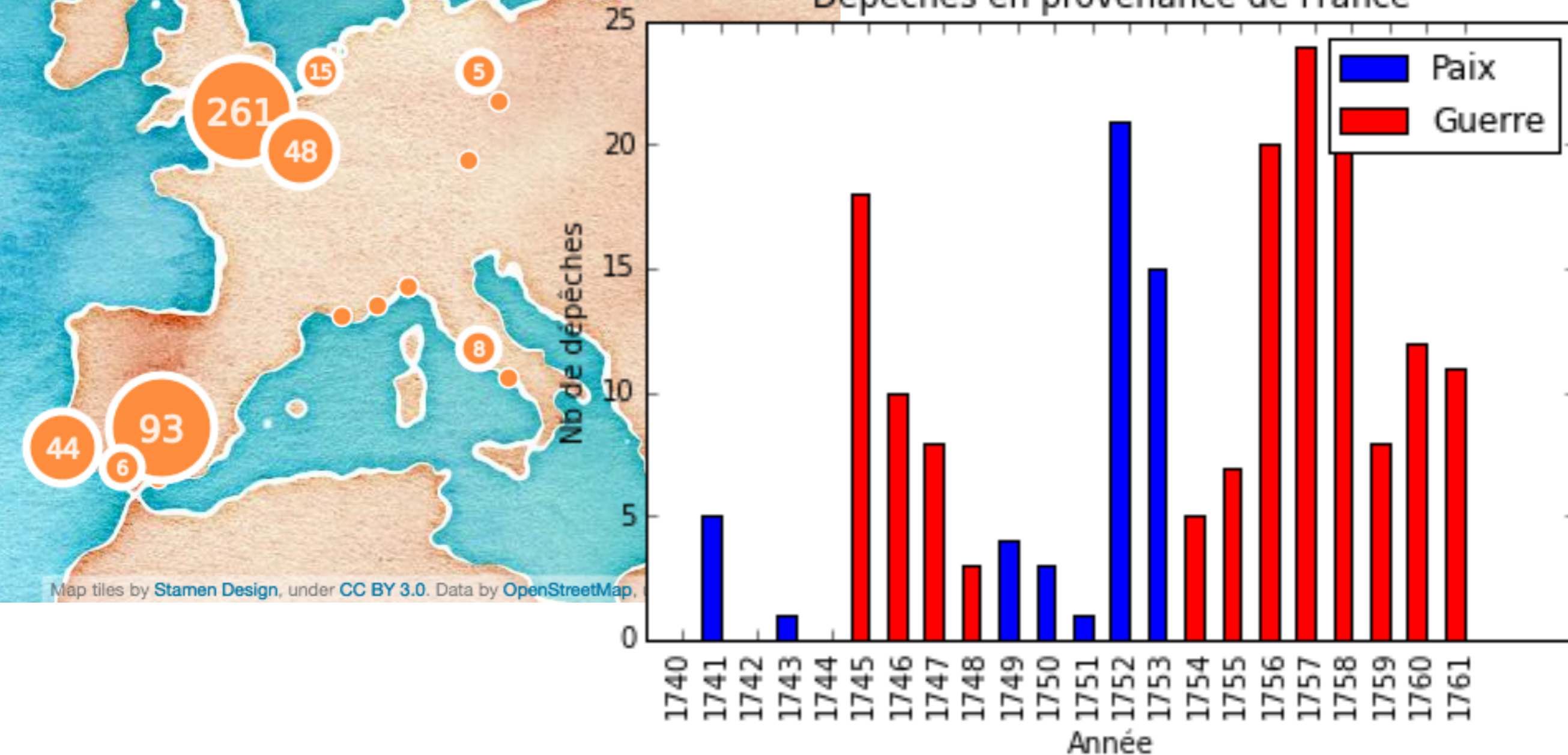
GAZETTE – RÉPARTITION GÉOGRAPHIQUE DES ORIGINES



GAZETTE – RÉPARTITION TEMPORELLE DES ARTICLES

Provenance des dépêches (1740-43 et 1749-53)

Dépêches en provenance de France



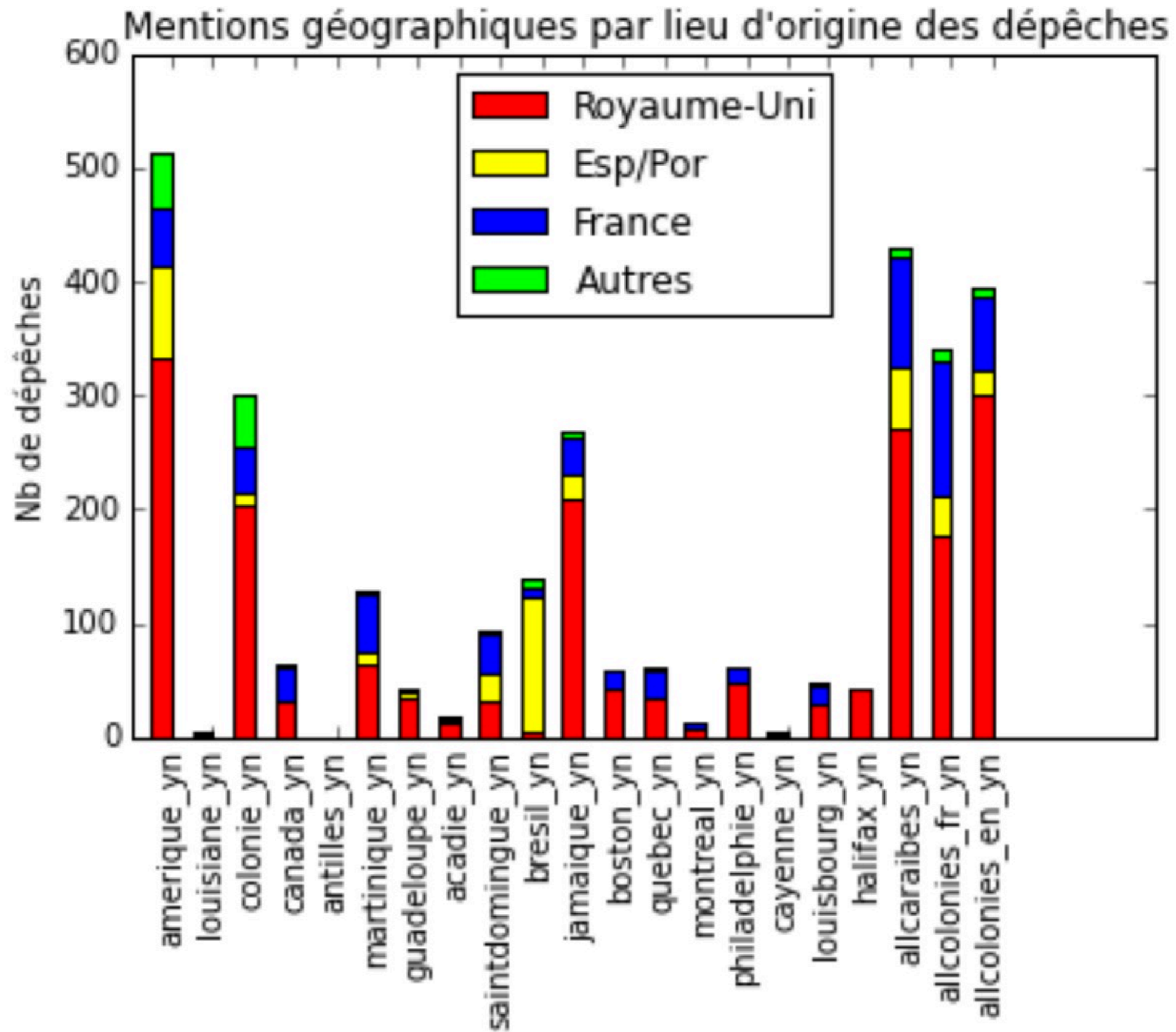
COOCCURRENCES DE LA FORME « CANADA »

- Par indice de spécificité : Canada, Sauvages, Fort, Vaudreuil, Lac, Détachement, Montcalm, Marquis, Troupes
- Par décompte absolu : Fort, Troupes, Canada, Marquis, Hommes, Sauvages, Anglois, Corps, Général, Ennemis, Détachement

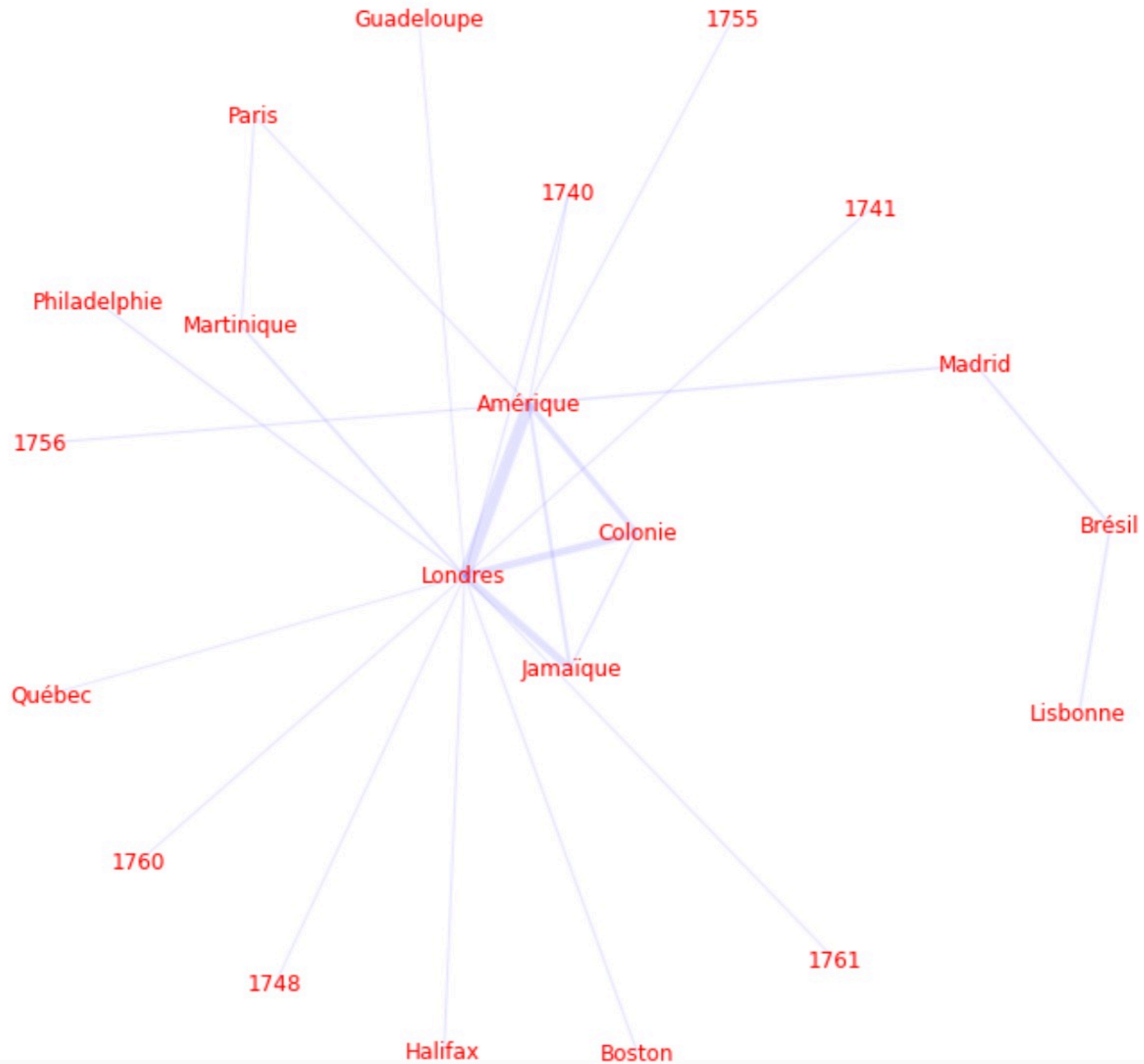
COOCCURRENCES DE LA FORME « MARTINIQUE »

- Par indice de spécificité : Capitaine, Martinique, Armateur, François, Rançonné, Navires, Prises, Pris, Vaisseau, Sucre, Enlevé, Chargé
- À +/- 10 formes de distance : Saint-Domingue, Navire(s), Conduit

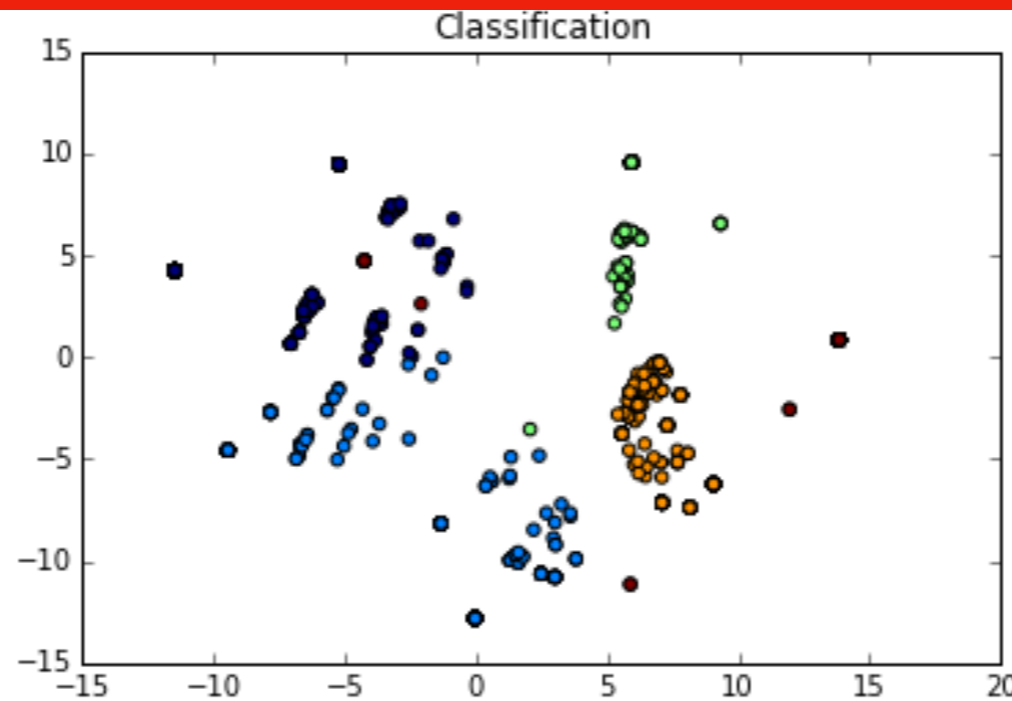
GAZETTE — ORIGINES DES ARTICLES MENTIONNANT CHAQUE MOT-CLÉ



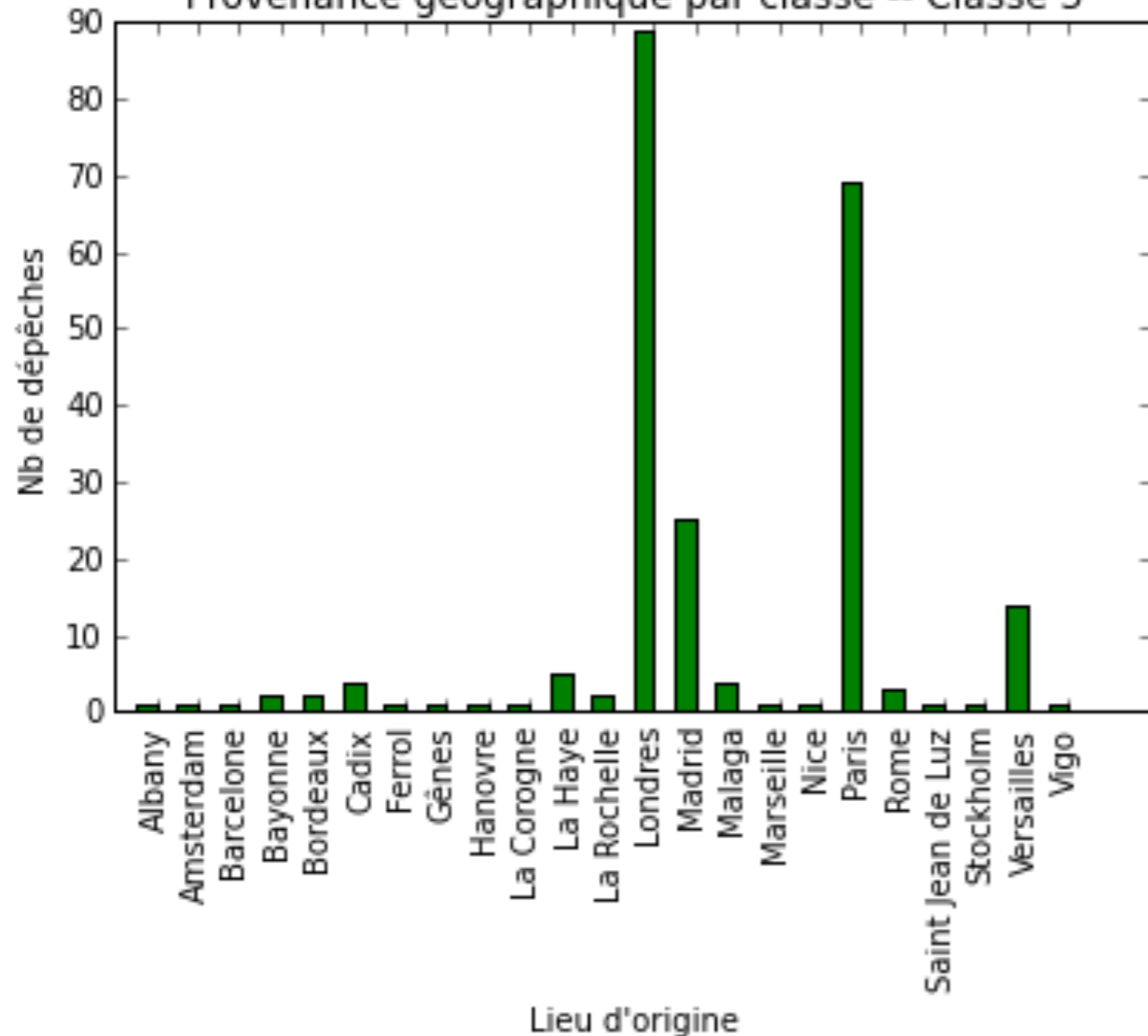
L'AMÉRIQUE DANS LA GAZETTE – COPRÉSENCE DE CONCEPTS



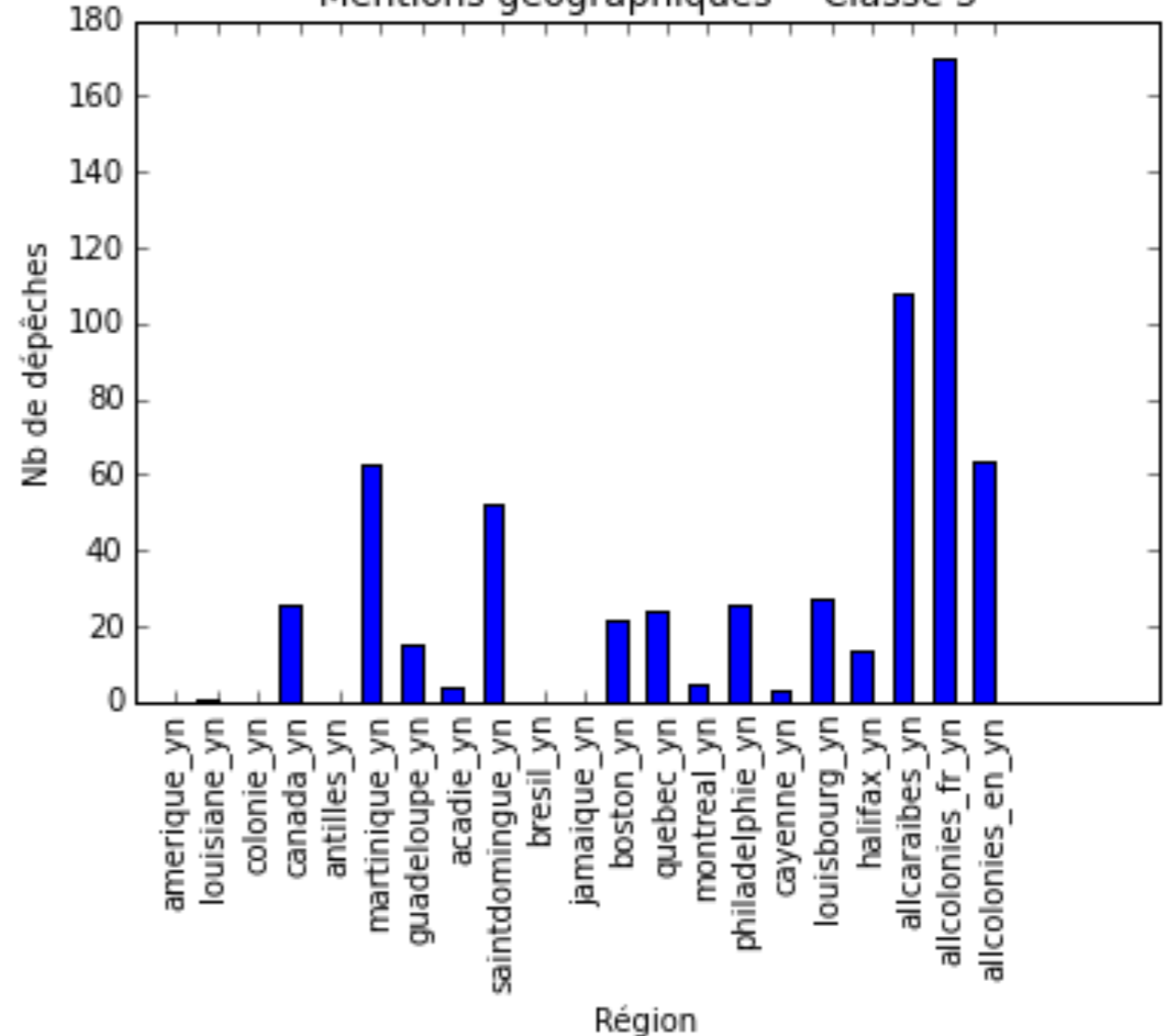
L'AMÉRIQUE DANS LA GAZETTE – CLASSIFICATION K-MOYENNES



Provenance géographique par classe -- Classe 3



Mentions géographiques -- Classe 3



TRAITER UN CORPUS TRANSCRIT : L'ENCYCLOPÉDIE

Diderot

1

ENCYCLOPÉDIE,
OU
DICTIONNAIRE RAISONNÉ
DES SCIENCES,
DES ARTS ET DES MÉTIERS,
RECUEILLI
DES MEILLEURS AUTEURS
ET PARTICULIÈREMENT
DES DICTIONNAIRES ANGLAIS
DE CHAMBERS, D'HARRIS, DE DYCHE, &c.
PAR UNE SOCIÉTÉ DE GENS DE LETTRES.
Mis en ordre & publié par M. *DIDEROT*; & quant à la PARTIE MATHÉMATIQUE,
par M. *D'ALEMBERT*, de l'Académie Royale des Sciences de Paris
& de l'Académie Royale de Berlin.
*Tantum series juncturae pollet,
Tantum de medio sumptis accedit honoris! HORAT.*
DIX VOLUMES IN-FOLIO,
DONT DEUX DE PLANCHES EN TAILLE-DOUCE,
PROPOSÉS PAR SOUSCRIPTION.

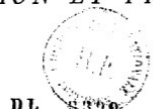


559

2

A PARIS, Chez {
BRIASSON, rue Saint Jacques, à la Science.
DAVID l'aîné, rue Saint Jacques, à la Plume d'or.
LE BRETON, Imprimeur ordinaire du Roy, rue de la Harpe.
DURAND, rue Saint Jacques, à Saint Landry, et au Griffon.

M. DCC. LI.
AVEC APPROBATION ET PRIVILEGE DU ROY.



DL 8320 15-3-51 A

La géographie dans l'*Encyclopédie*

- ~20% de tous les articles
- Contenu transversal

L'*Encyclopédie* dans la société européenne

- Le « best-seller » du siècle
- Un colosse économique

LA GÉOGRAPHIE DANS L'ENCYCLOPÉDIE

ARTFL Encyclopédie Project - Robert Morrissey, General Editor; Glenn Roe, Assoc. Editor

[Browse](#)

[Search](#)

[Advanced Search](#)

To look up a word in a dictionary, select the word with your mouse and press 'd' on your keyboard.

Searching for **normal class=géographie** in the whole database. Found 14547 objects.

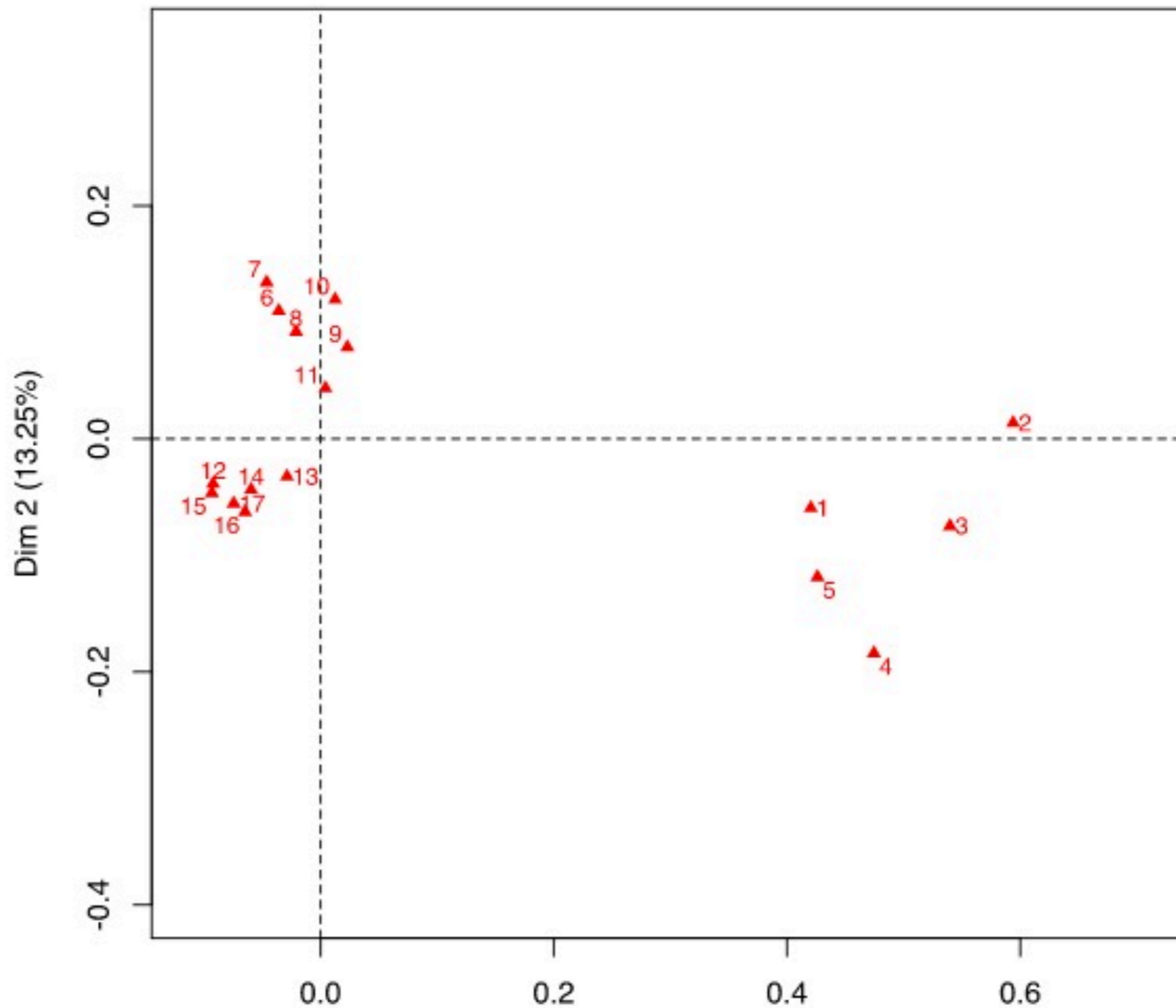
Abyde, (Geog. anc.) [Géographie ancienne] Diderot
ACÉ, (Geog. anc.) [Géographie ancienne] unknown
ACHAIE, (Geog. anc.) [Géographie ancienne] Diderot
ACHERON, (Geog. anc. & Myth.) [Géographie ancienne , Mythologie] Diderot
ACHERUSE, (Geog. Hist. anc. & Myth.) [Géographie , Histoire ancienne , Mythologie] Diderot
ACHILLEA, (Geog. anc.) [Géographie ancienne] Diderot
ACQS, (Geog.) [Géographie] Diderot
ACRE, (Geogr.) [Géographie] Diderot
ACROCERAUNES, (Geog. anc. & mod.) [Géographie moderne , Géographie ancienne] Diderot
ADA, (Geog. mod.) [Géographie moderne] Diderot
ADAMA, (Geog. anc.) [Géographie ancienne] Diderot

Corpus géographique complet

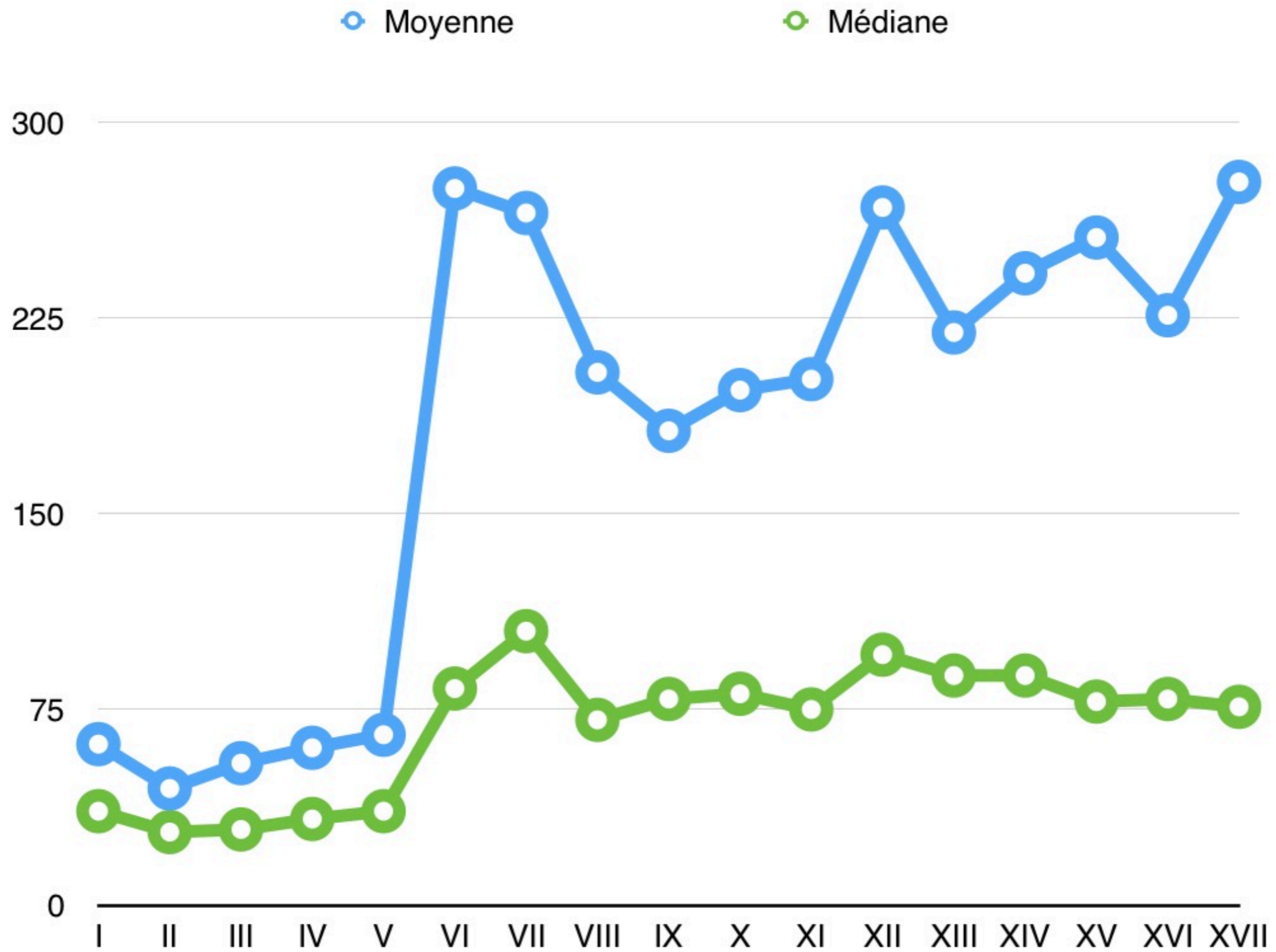
- 14 547 articles
- 5 250 articles de géographie moderne
- 3 457 articles de géographie ancienne
- 5 800 articles de géographie (générale)
- ~ 200 articles de catégories mineures
- Quelques articles partagés

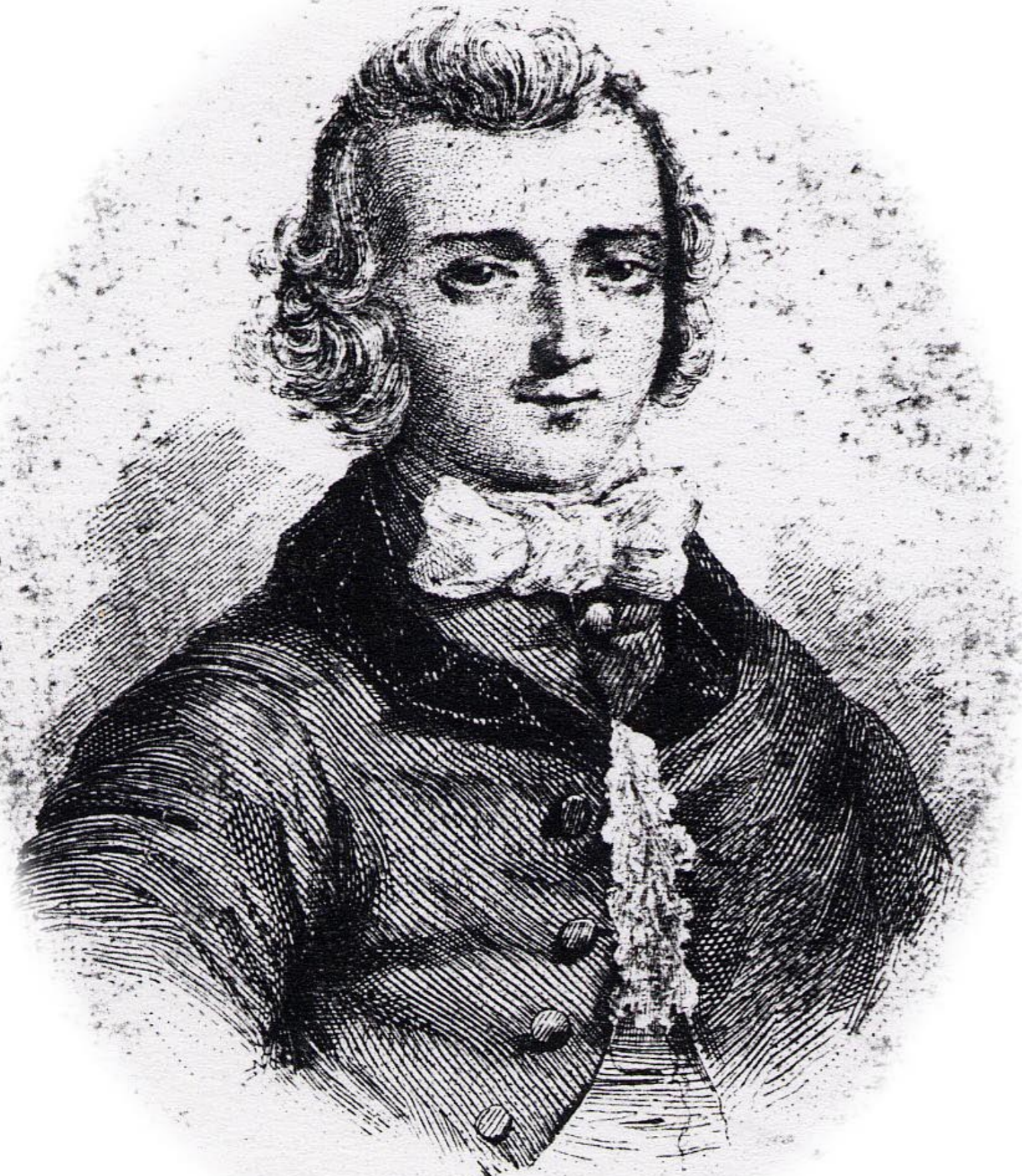
ANALYSE FACTORIELLE DU CORPUS DIVISÉ PAR VOLUME

Plan factoriel de l'analyse des correspondances
sur la partition Sans 9489, par volume du corpus Sans article 9489



EXPLORATION DES MÉTADONNÉES





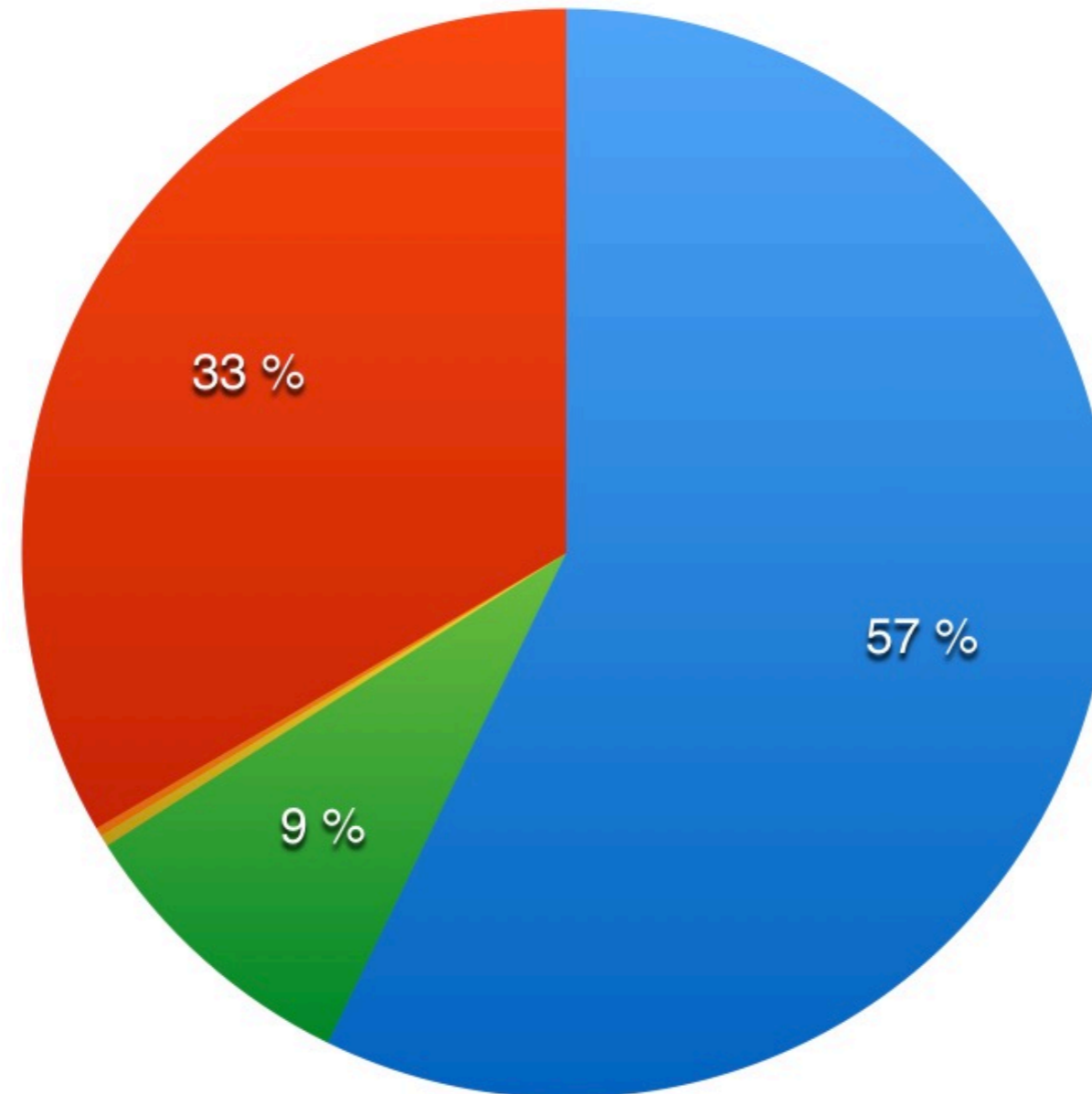
Louis de Jaucourt

*Auteur de ~17 400
articles, dont 8 323
articles géographiques.*

[Image: Wikimedia Commons](#)

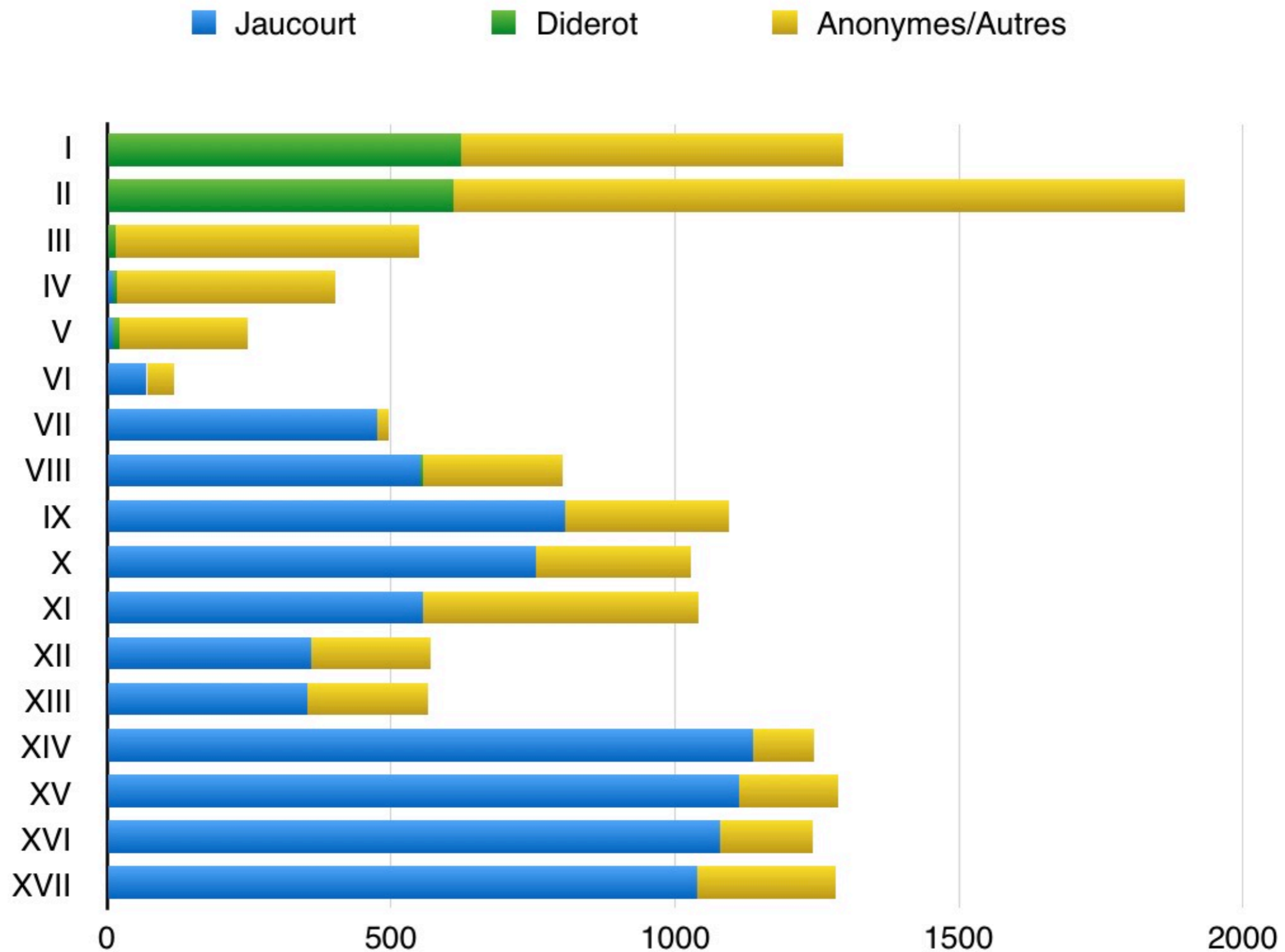
EXPLORATION DES MÉTADONNÉES (suite)

● Jaucourt ● Diderot ● D'Alembert ● Autres ● Anonymes



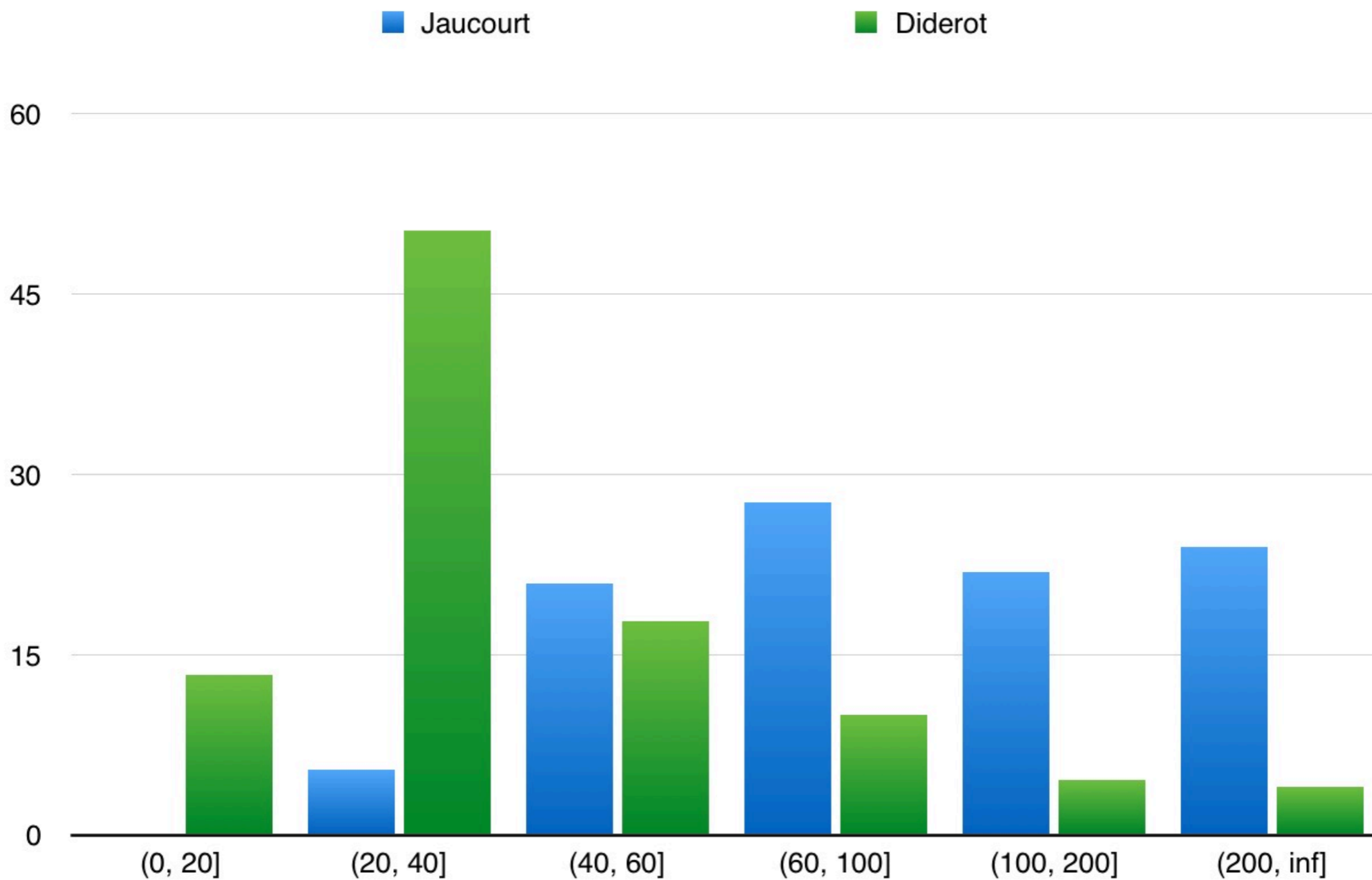
Répartition du corpus selon les auteurs

EXPLORATION DES MÉTADONNÉES (suite)

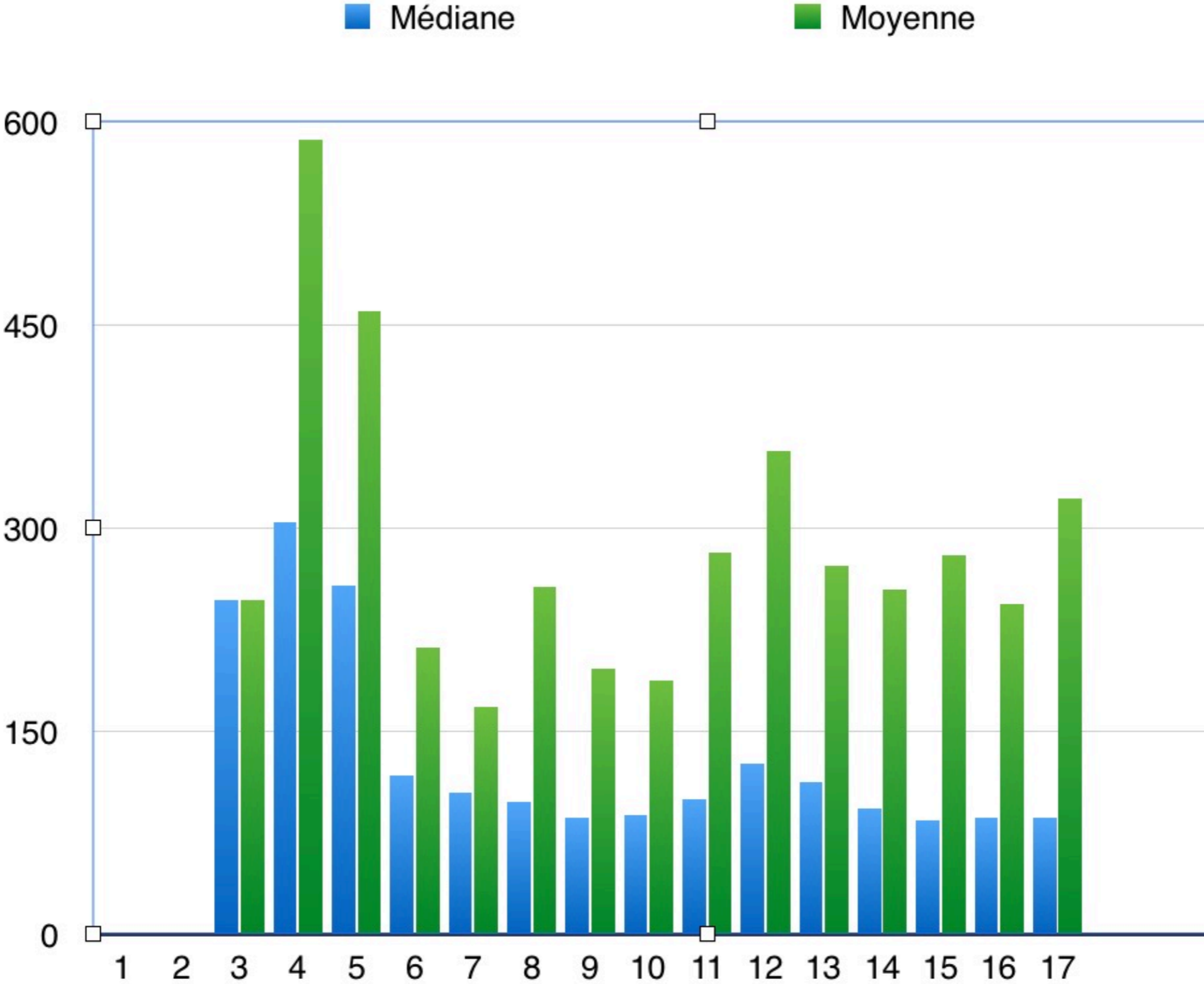


Articles par auteur, par volume

DISTRIBUTION DES LONGUEURS D'ARTICLES PAR AUTEUR



LONGUEUR DES ARTICLES DE JAUCOURT, PAR VOLUME



RICHESSSE DU VOCABULAIRE PAR AUTEUR

Auteur	Formes uniques	Formes utilisées 5x ou moins	Hapax
Diderot	8 122 (6,38 p.a.)	7 442 (5,84 p.a.)	5 481 (4,30 p.a.)
Anonymes/ Autres	30 260 (6,11 p.a.)	26 032 (5,26 p.a.)	17 761 (3,59 p.a.)
Jaucourt	78 685 (9,45 p.a.)	64 477 (7,75 p.a.)	41 669 (5,01 p.a.)

Vocabulaire (p.a. = par article)

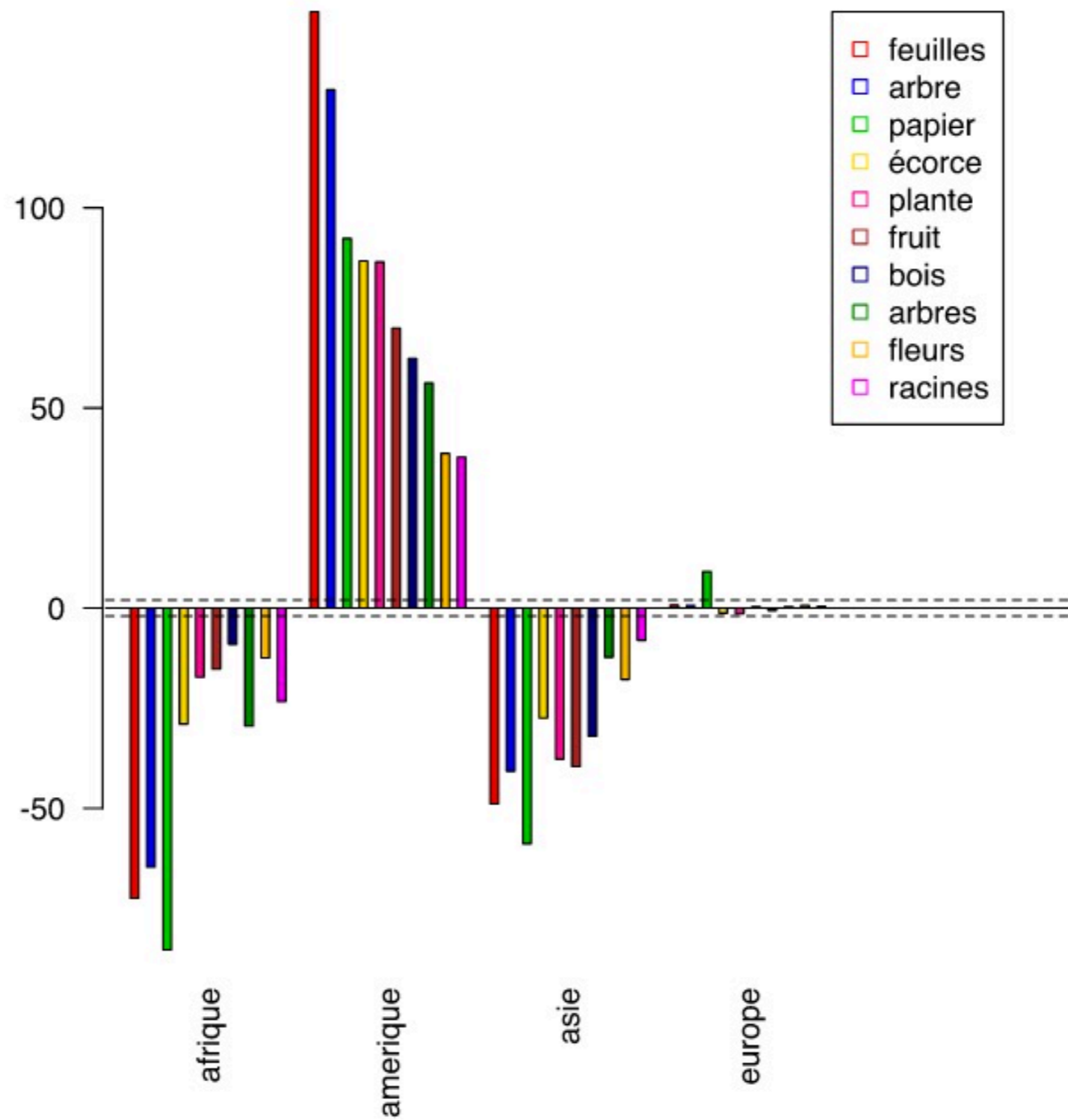
UN PROJET SCIENTIFIQUE DÉTOURNÉ?

Le chevalier de Jaucourt ? — Ne craignez pas qu'il s'ennuie de moudre des articles ; Dieu le fit pour cela. Je voudrais que vous vissiez comme sa physionomie s'allonge quand on lui annonce la fin de son travail, ou plutôt la nécessité de le finir. Il a vraiment l'air désolé.

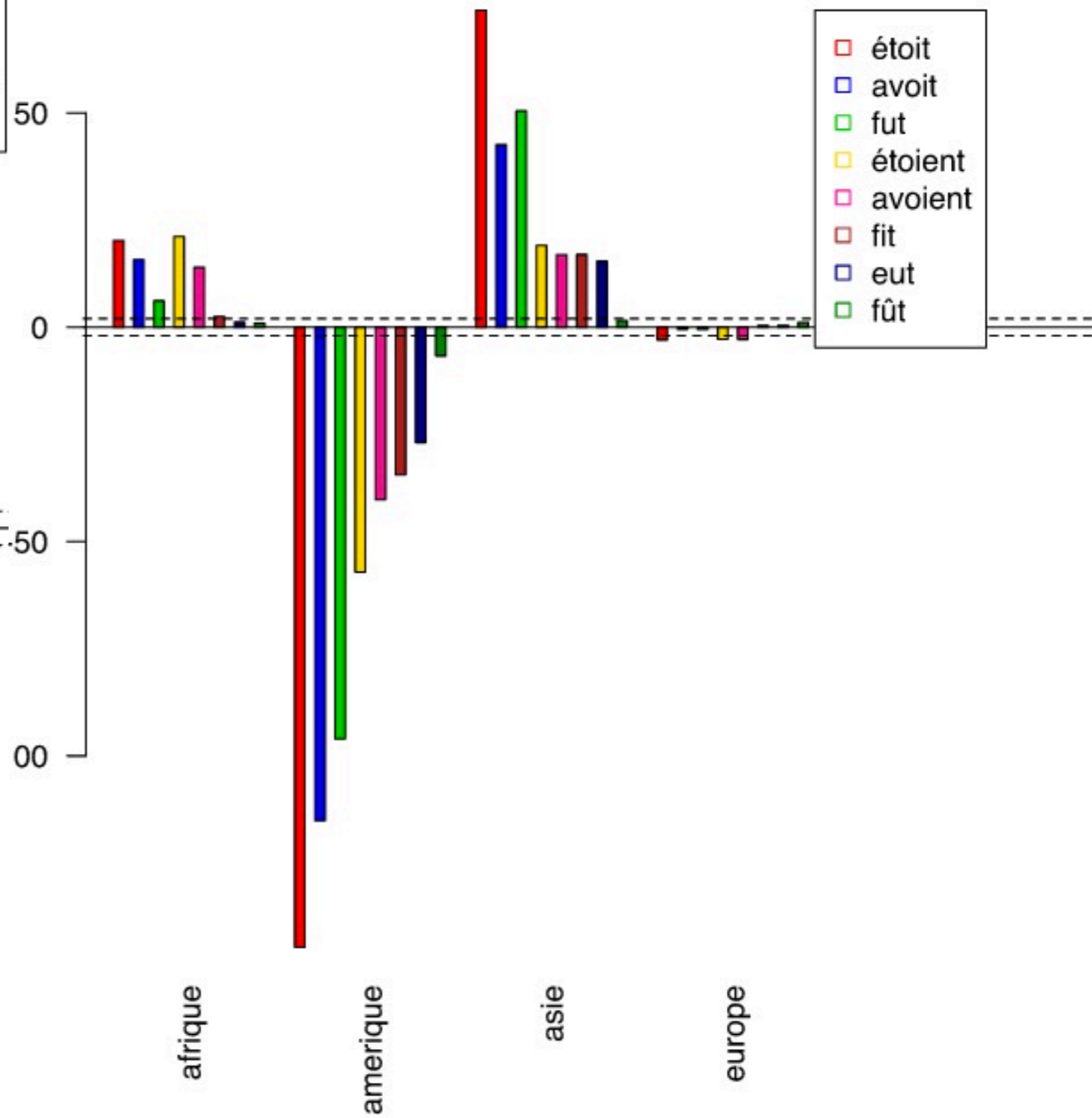
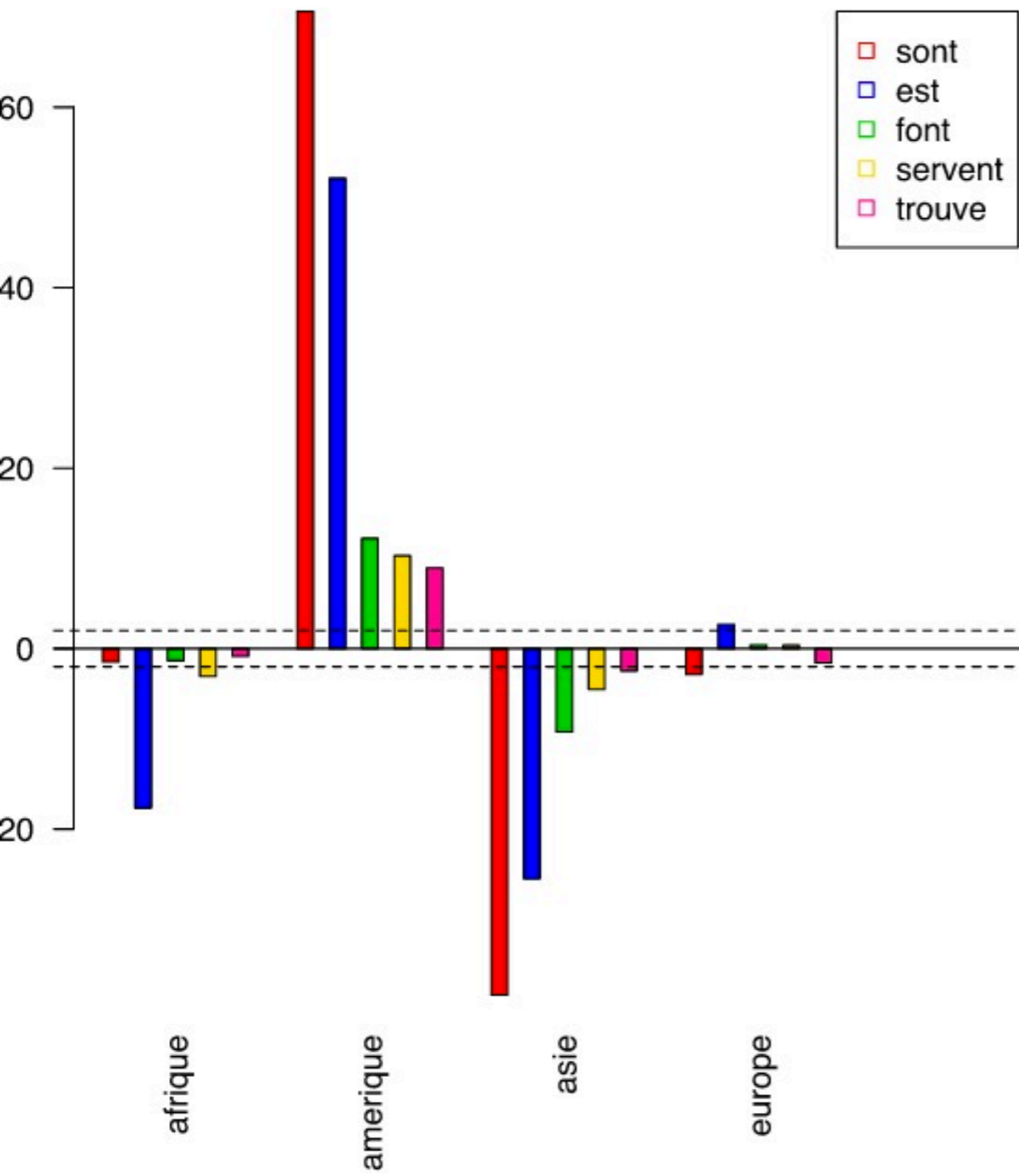
— *Lettre de Denis Diderot à Sophie Volland, 25/11/1760*

Source: Diderot, Oeuvres, Laffont, t. V, p. 326

SPÉCIFICITÉS LEXICALES



SPÉCIFICITÉS LEXICALES (suite)



CONCLUSIONS

- **Un outil utile...**
- **Mais un outil seulement**
- **Compromis entre ampleur et profondeur**
- **Interprétation demeure la chasse gardée de l'humain**

Merci!

Courriel: flaramee@uottawa.ca

Web: www.francoisdominiclaramee.com

Twitter: @fdlaramee

ORCID: 0000-0001-5542-3754