

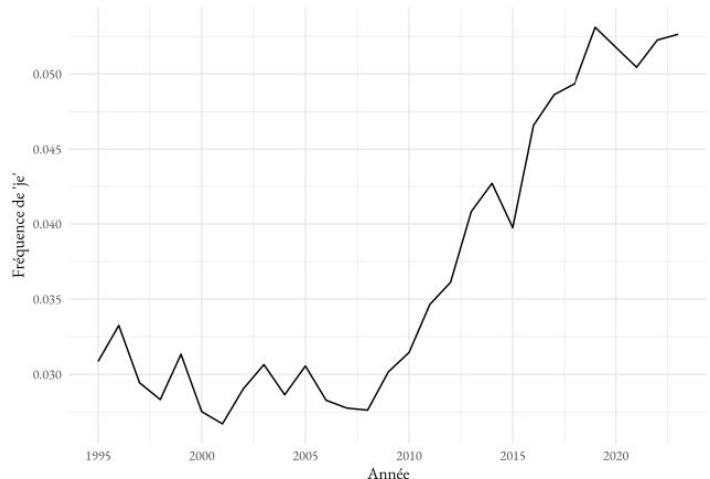
Une approche lexicométrique du rap français

Juliette Hubert (Université Laval & Université Polytechnique Hauts-de-France)
featuring

Benoît de Courson (École Normale Supérieure-PSL & Leiden University)

Introduction

- Au commencement était Gallicagram
- “Ma qualité de lyrics est quantifiable”
(Médine, 2018)
- Juliette, une alliée de taille



Juin :

- * 01/06 : Årsenik - Quelques gouttes suffisent...
- * 03/06 : Ad'Hoc-1 - À Bord Du N-333
- 05/06 : Kheops - Sad Hill (réédition 1CD)
- * 15/06 : Dee Nasty - Le diamant est éternel
- * 15/06 : Saïan Supa Crew - Saïan Supa Land (EP)
- * 19/06 : Basic (Vestat & Endo) - 98 jusqu'à l'infini...
- * 19/06 : Mellowman - Au jour le jour
- * 22/06 : Le Péril Noir - Le destin
- * 23/06 : I.m.s. - La bible du soldat

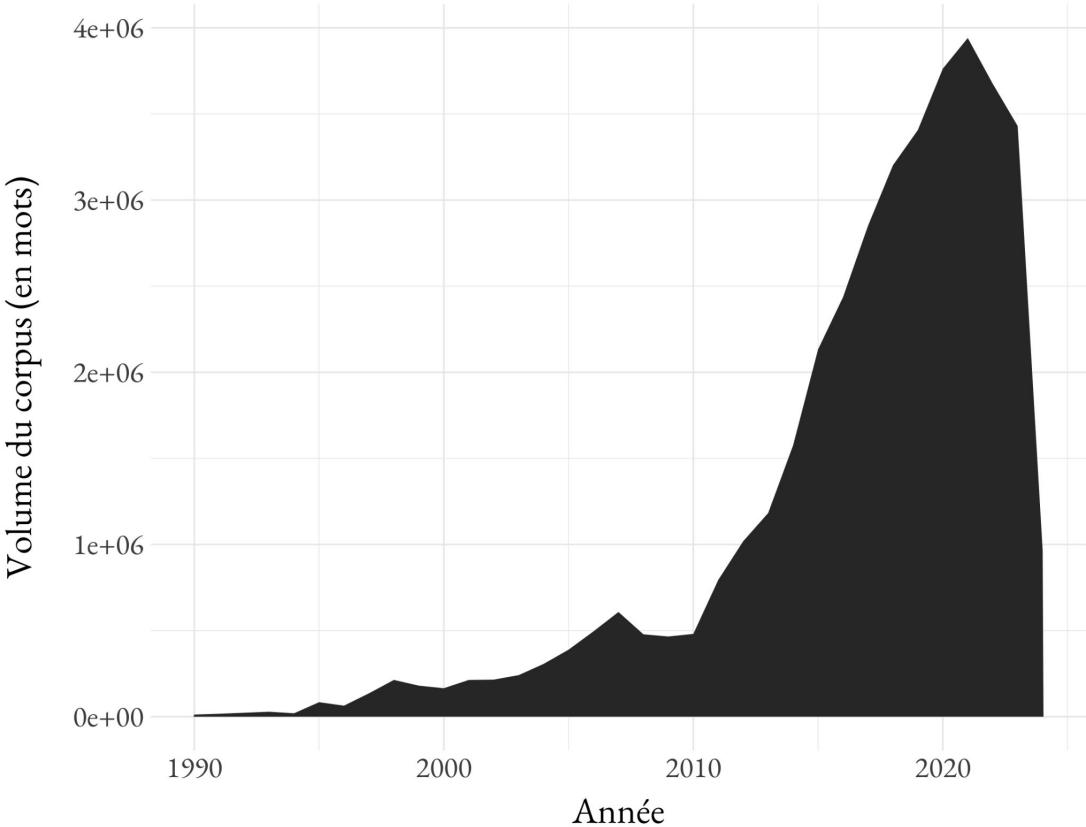
Juillet :

- 01/07 : Kabal - États d'Âmes
- 21/07 : MC Solaar - MC Solaar

Edit Lyrics

Cette année, j'fête mes dix ans au Wakanda
Ma qualité d'lyrics est quantifiable
Ne confonds pas Paul Bocuse et l'McDonald's
Tous tes rappeurs sont là par le piston
Moi, je ne mens pas, c'est la vérité qui s'trompe
J'leur déclare la guerre toutes les sept secondes
J'commande mes hommes et mes pizzas sur le même ton
J'ai fini ma sèche, j'suis en prise de masse
J'ai fait de mes seize une musique de masse
Rap hardcore, rappeurs qui s'clashent
Même entre nous, y a des rapports de classe

Le corpus



Remarques :

- Rap comme entité
- Corpus
- Textualité

Plan

Que faire de ces données textuelles ? Un tutoriel à travers démarches possibles :

- Une batterie de mesures
- Une démarche *data-driven*
- Tester une hypothèse pré-existante

I. Une approche aveugle : la “batterie de mesures”

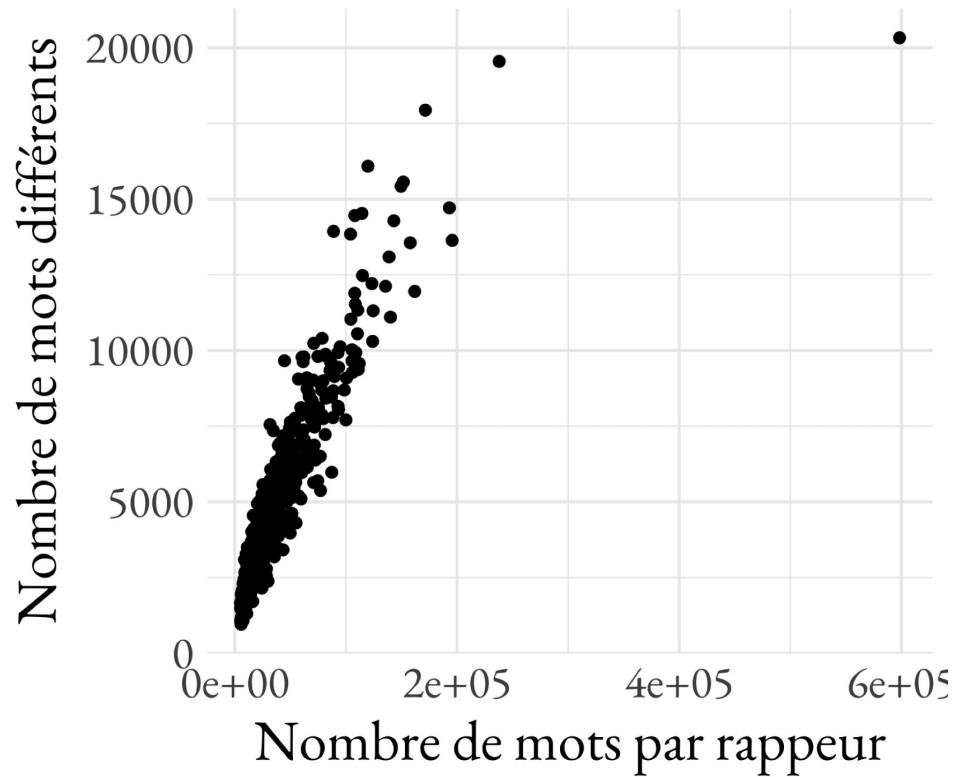
- Une ‘lecture distante’ peut appréhender un corpus de taille inhumaine
- On ne peut pas mesurer *la* qualité, mais on peut mesurer *des* qualités
- En particulier :
 - Les rappeurs sont plus ou moins « lyricistes », « obsédés textuels »
 - Les rappeurs sont plus ou moins violents verbalement
- Quel lien statistique entre les deux ? Quelle évolution ?

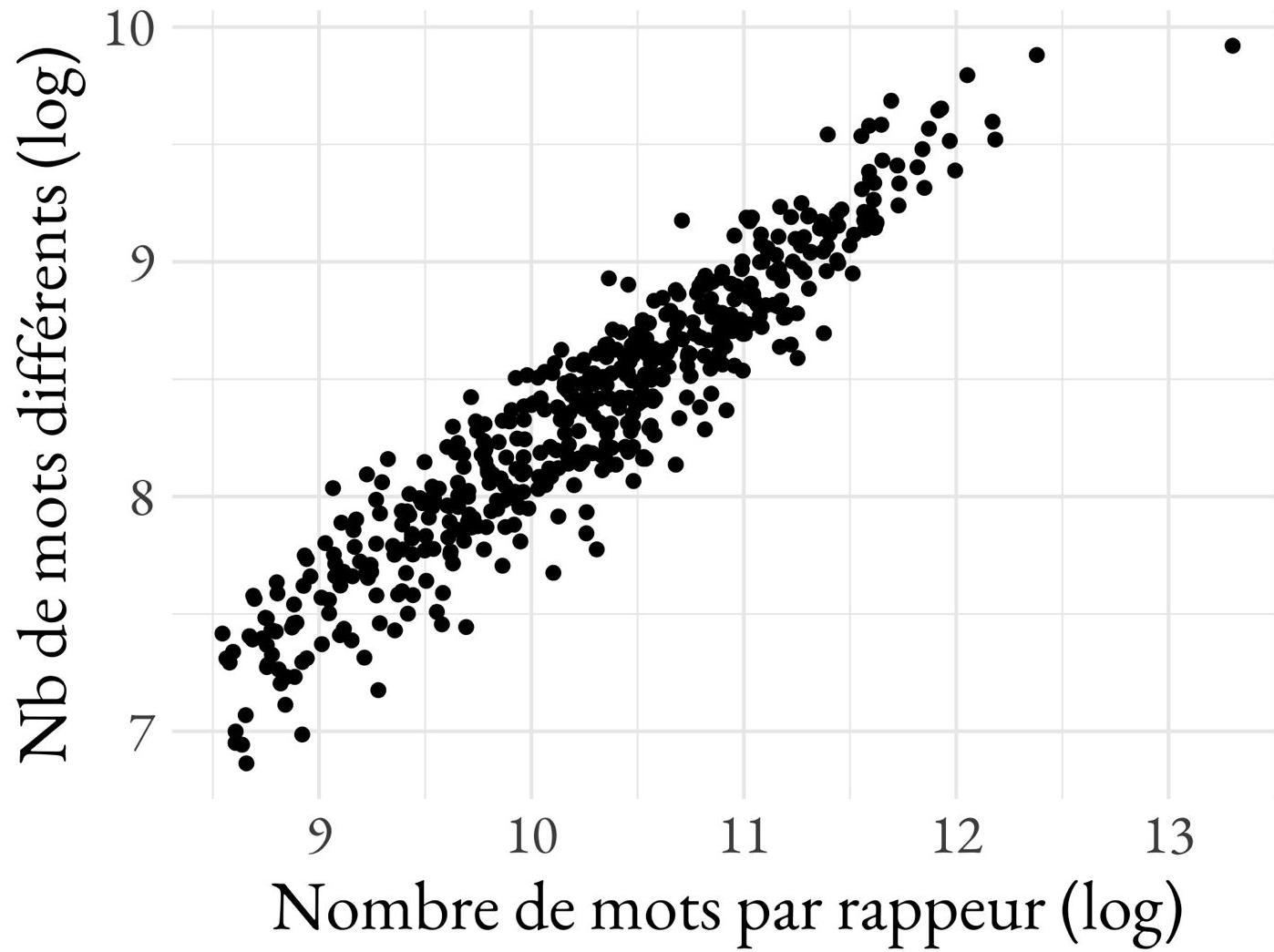
Que mesurer ?

- Des mesures de complexité (diversité lexicale, longueur des mots)
- La fréquence de lexiques (insultes, verlan, argot, ...)
- L'analyse de sentiments (LIWC)
- Appliquer des détecteurs de haine/sexisme

Comment mesurer l'étendue du vocabulaire ?

- On peut compter le nombre de mots différents
- ...mais JuL a publié près de 1000 morceaux, Tiakola seulement 24
- Et si on divisait le nombre de mots différents par le nombre de mots ?
- Le problème s'inverse





Nb de mots différents (log)

10

9

8

7

9

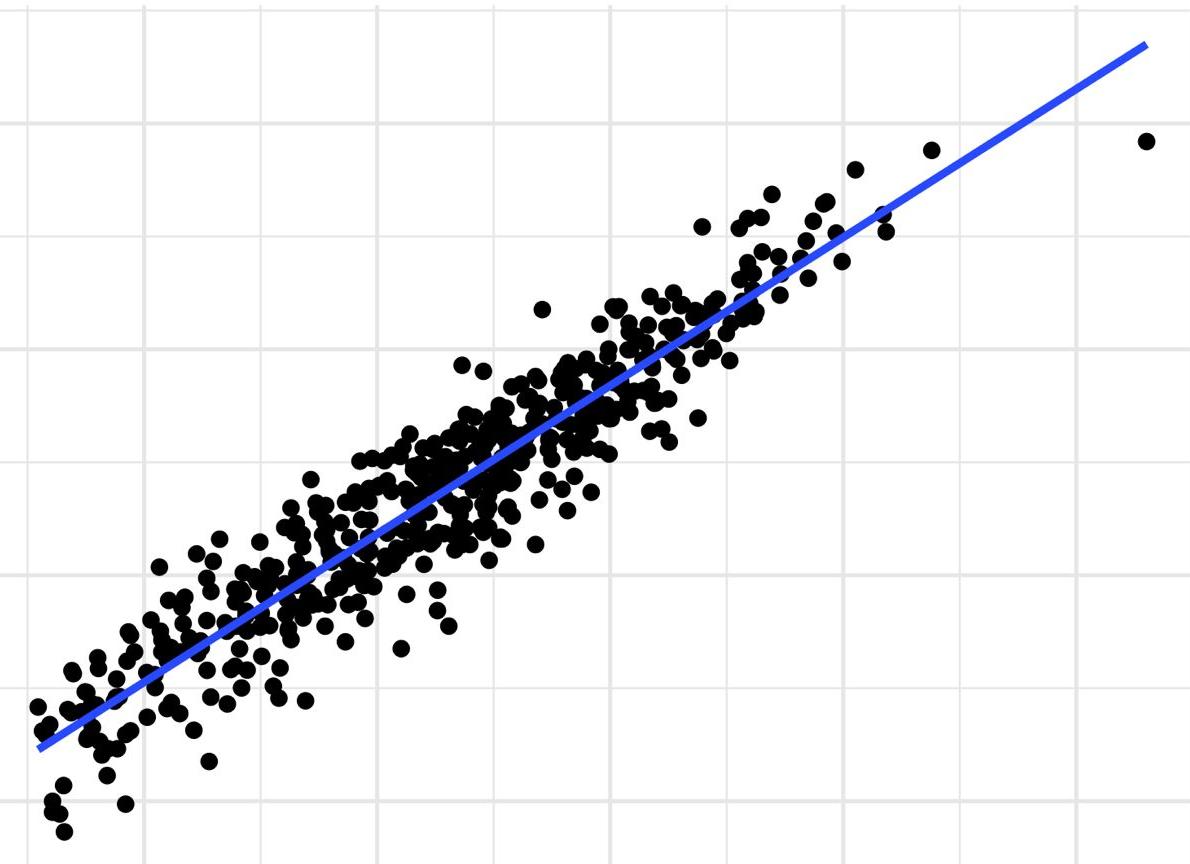
10

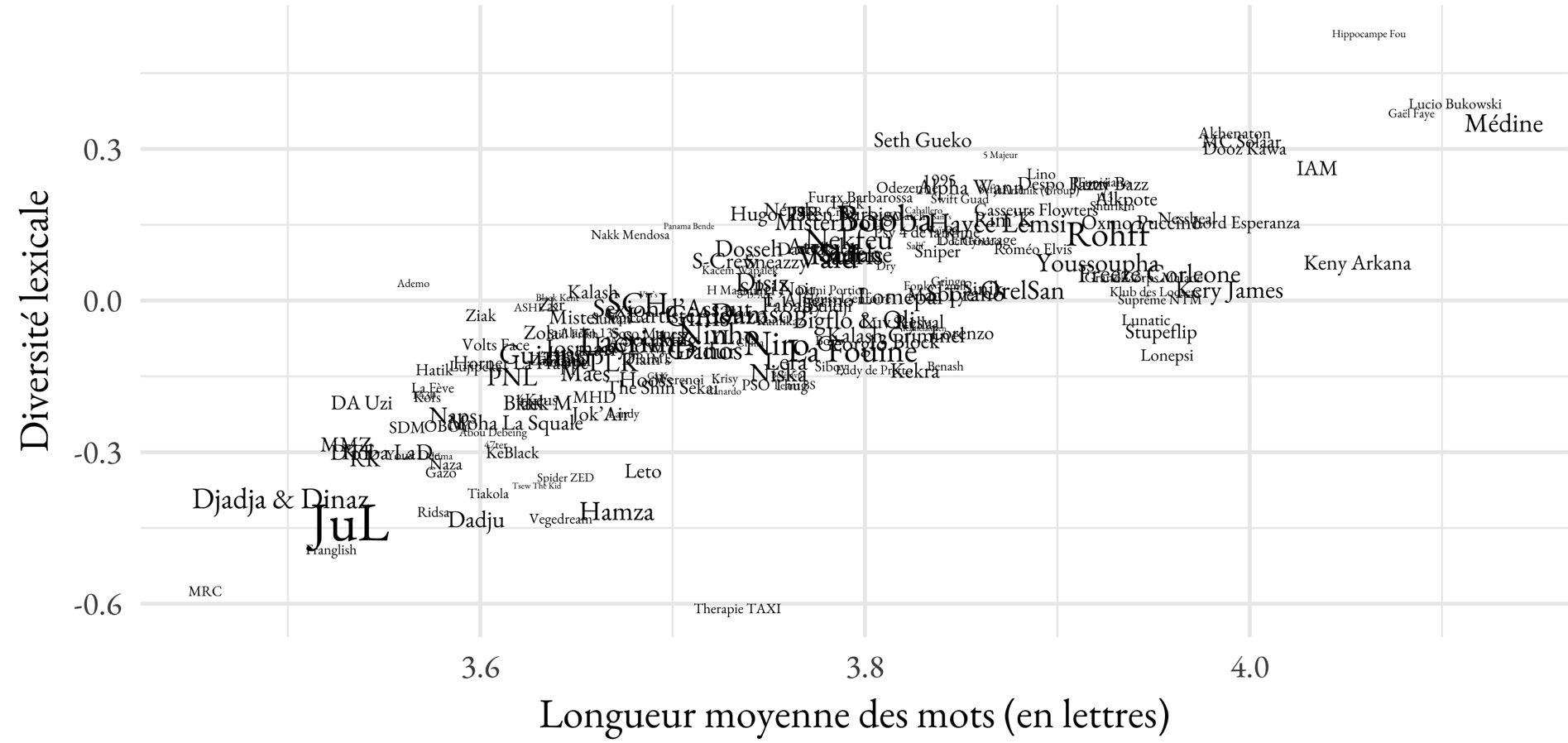
11

12

13

Nombre de mots par rappeur (log)





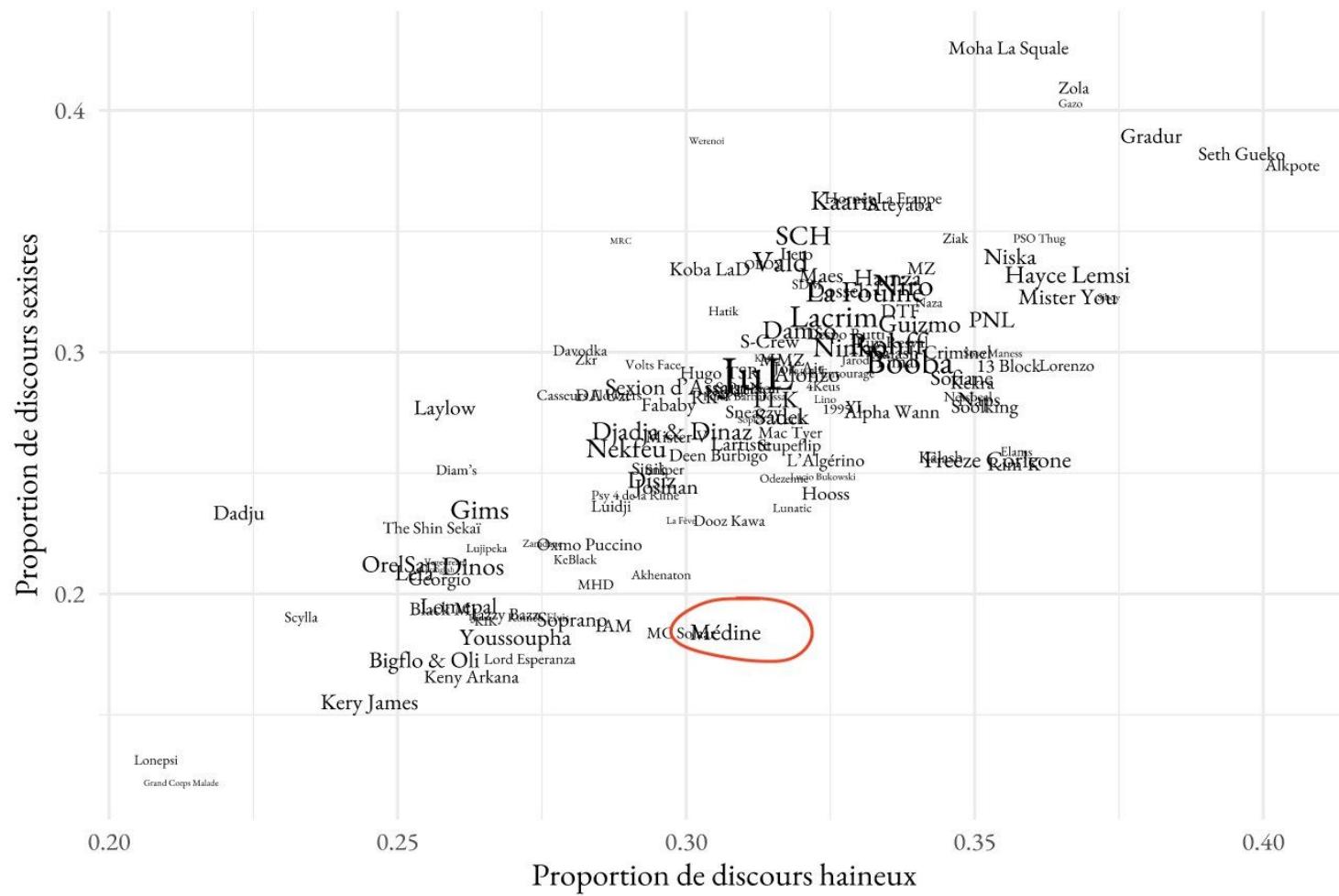
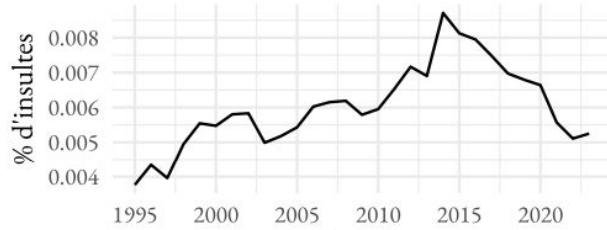
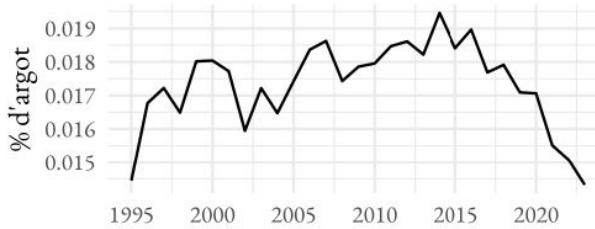
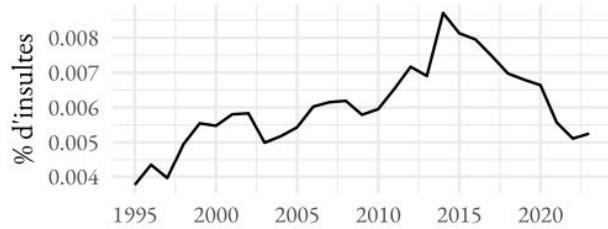


FIGURE 12 – Proportion de discours sexistes et haineux pour les rappeurs les plus vus

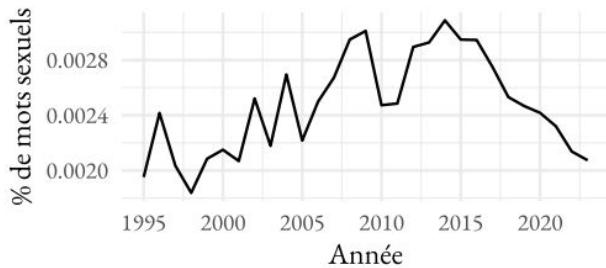
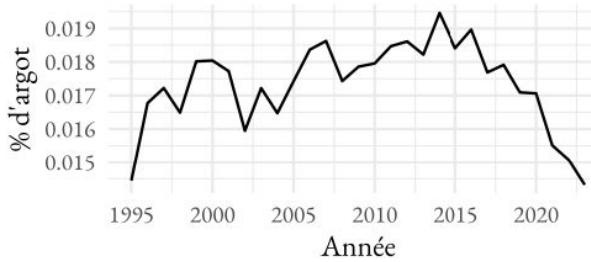
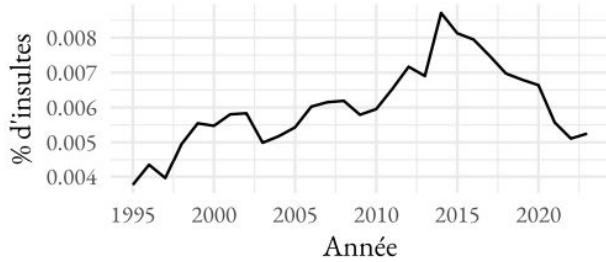
Une évolution en deux temps



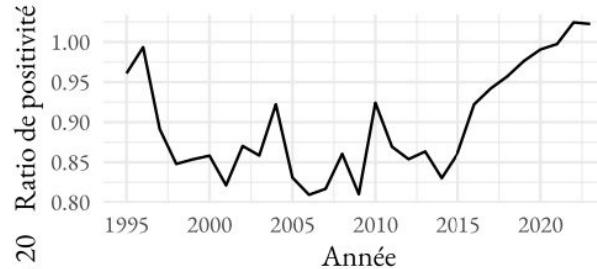
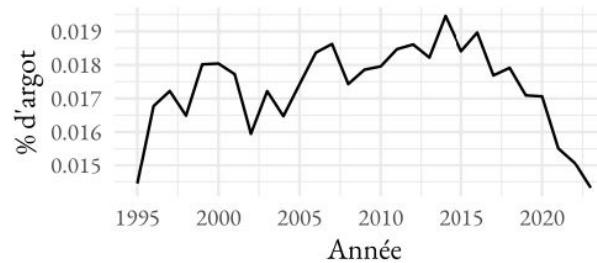
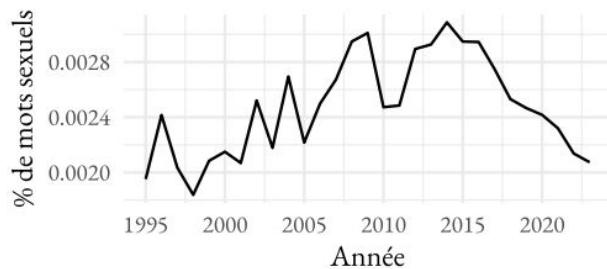
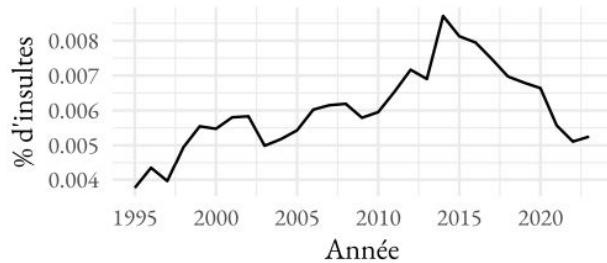
Une évolution en deux temps



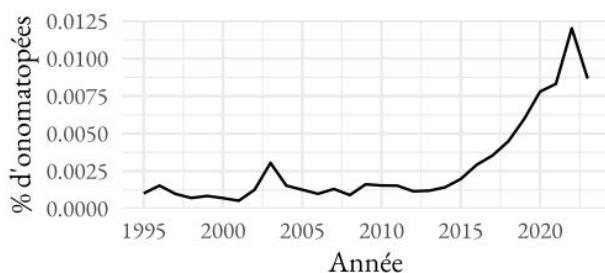
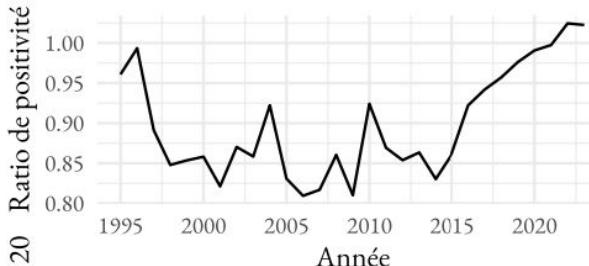
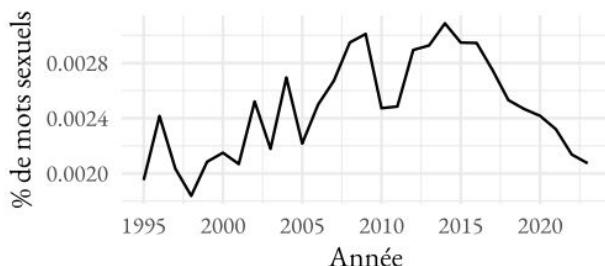
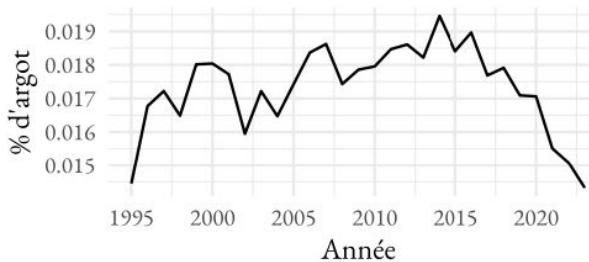
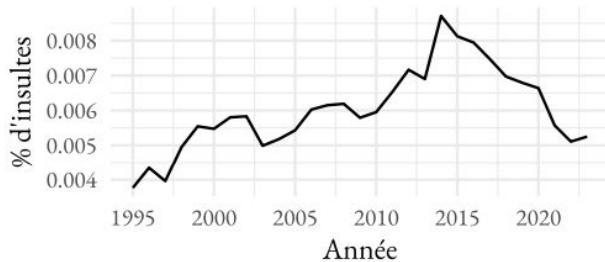
Une évolution en deux temps



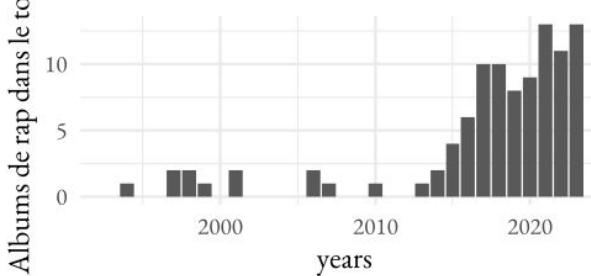
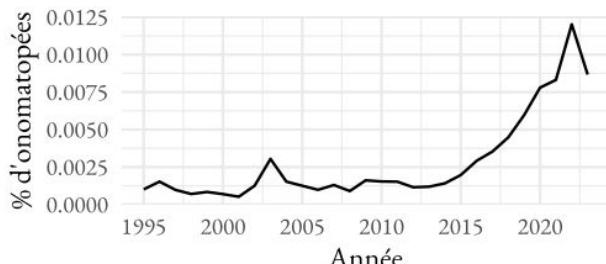
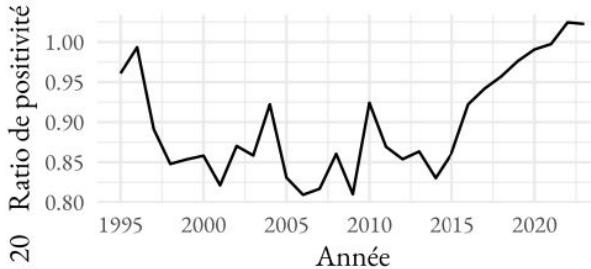
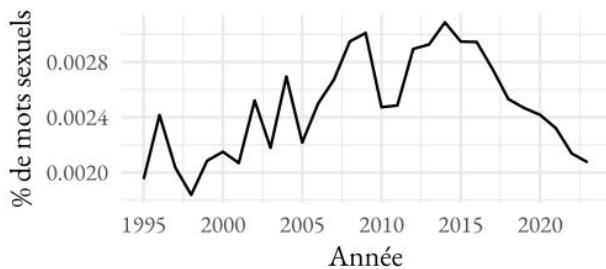
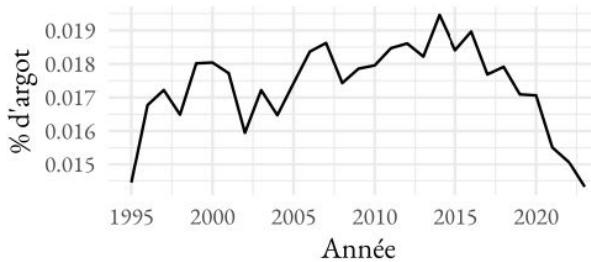
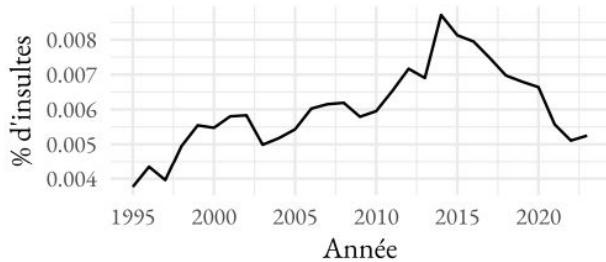
Une évolution en deux temps

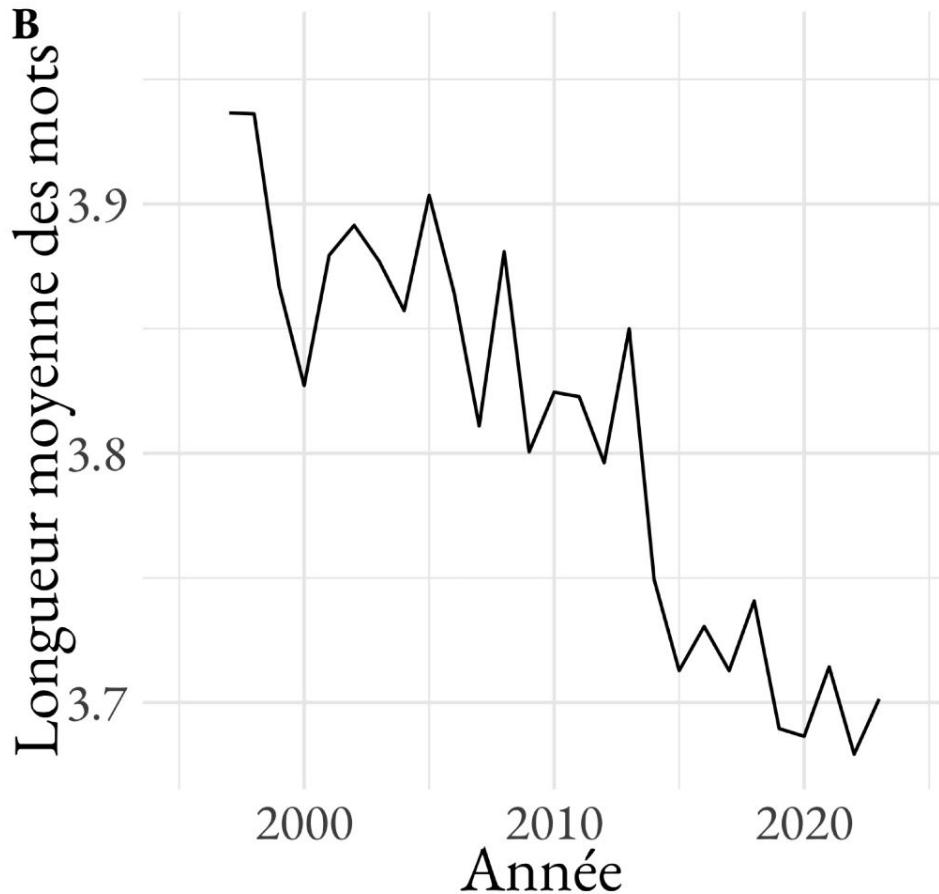
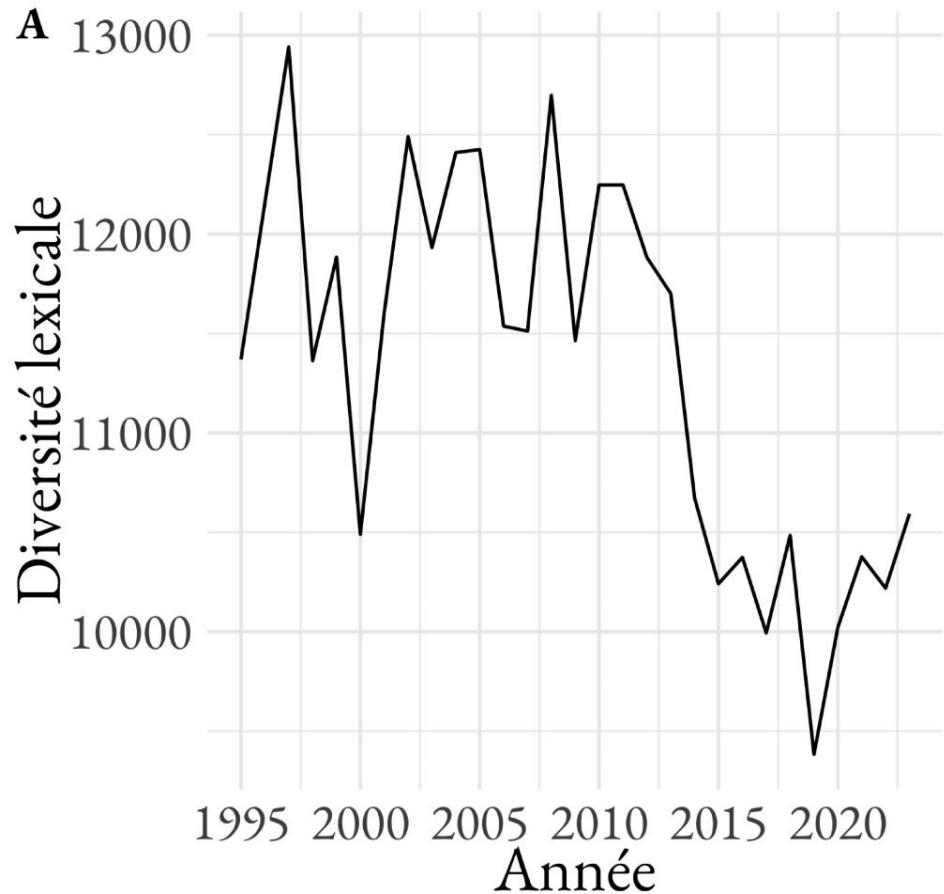


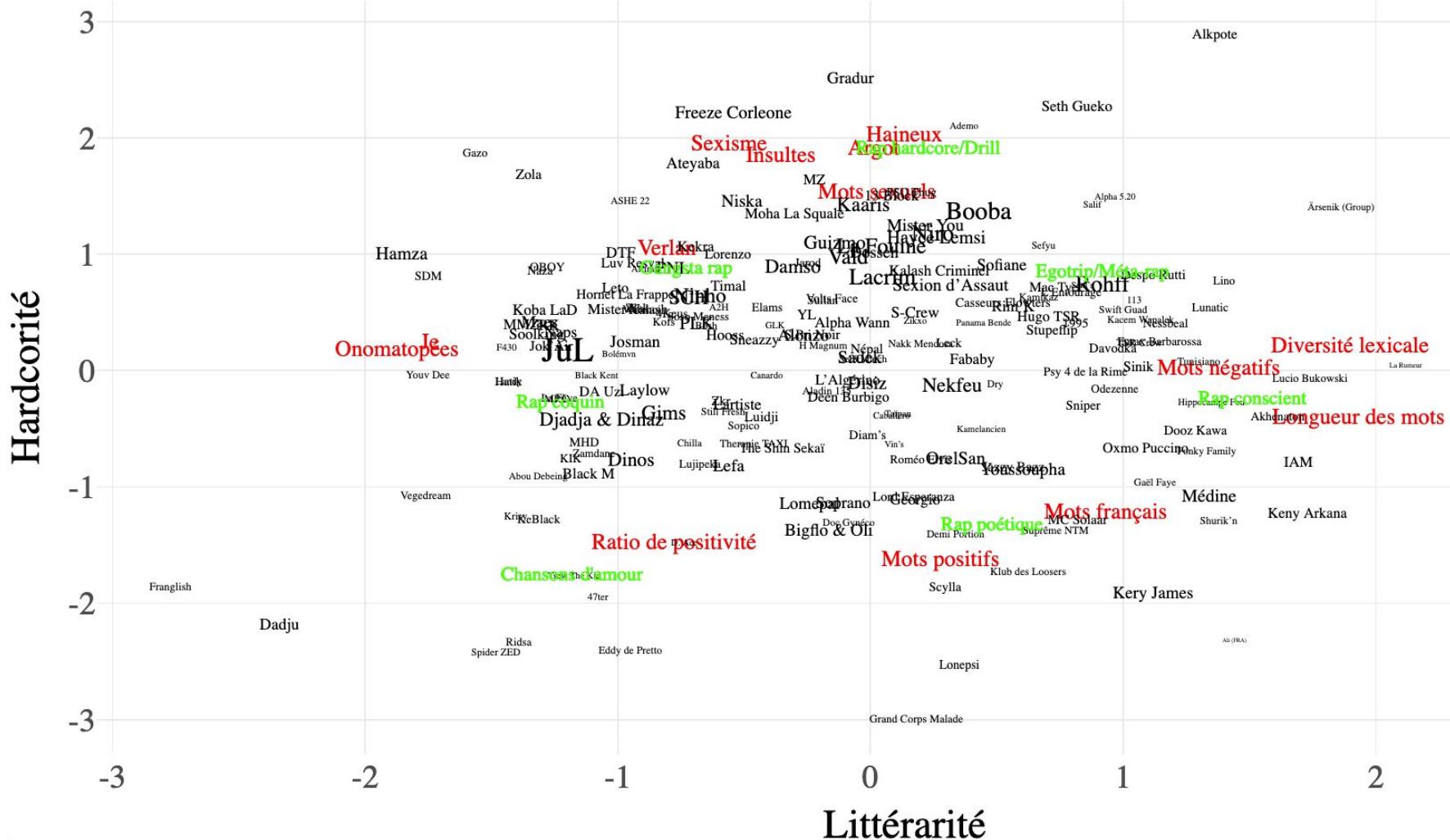
Une évolution en deux temps



Une évolution en deux temps







Où en est-on ? A-t-on objectivé le rap français ?

II - Une approche réellement data-driven ?

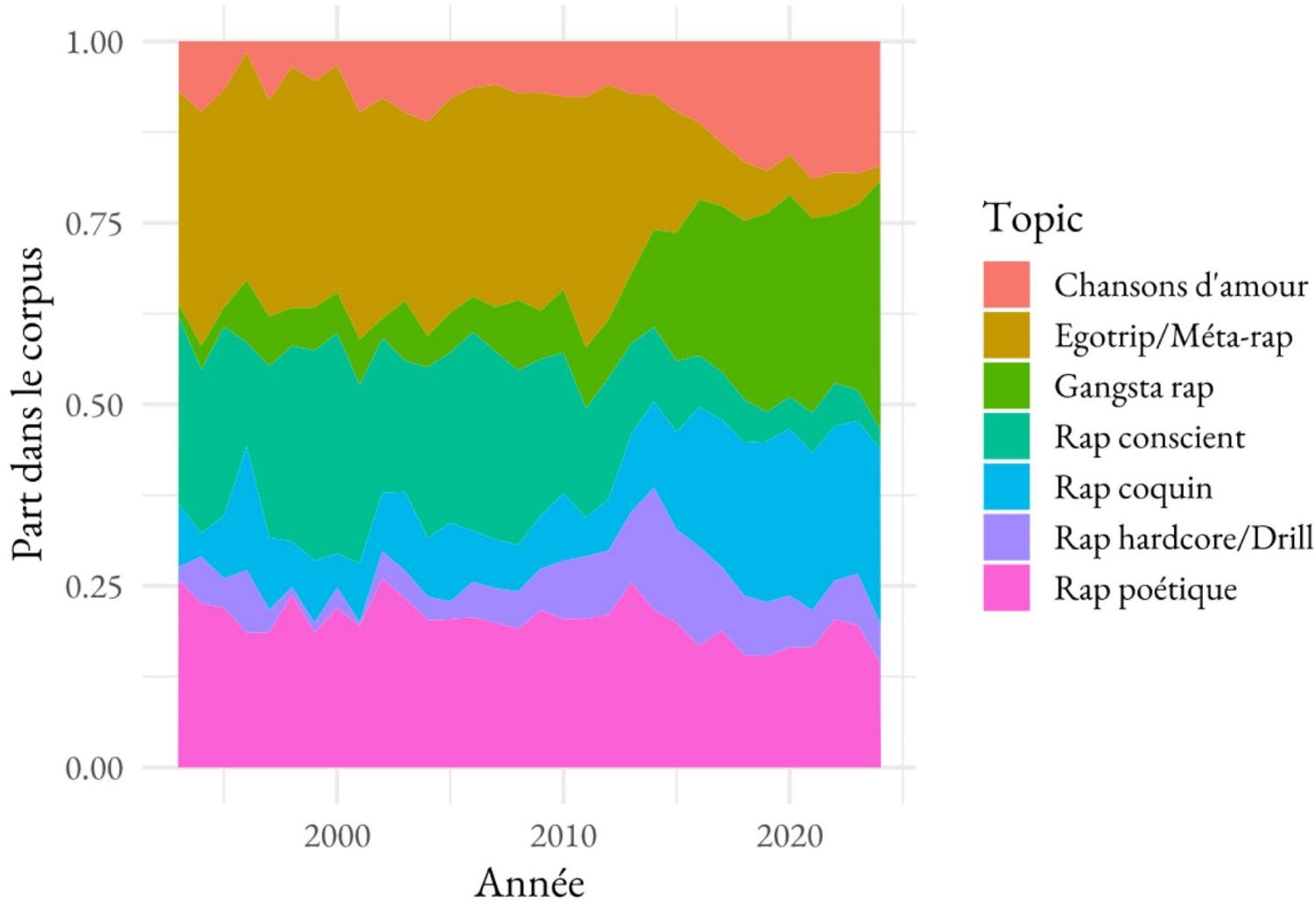
- Comment mesurer sans juger ?
- Le *topic modeling* : rassembler des textes sémantiquement similaires
- Une méthode “non-supervisée”

Le *topic modeling*

Mots caractérisant le topic	%	Nom (selon l'auteur)
beuh, sacoche, binks, billets, poto, monnaie, Audi, moula, zone, condés, euros, pétard, cli, mets, frappe rêves, ciel, espoir, bonheur, vie, larmes, lumière, souvenirs, jour, destin, âme, vide, monde, silence, soleil peuple, pays, système, justice, liberté, jeunes, violence, médias, familles, politique, État, droits, racisme, peuples, pauvres	19.96	Koba LaD
rap, mic, flow, MC, rimes, beat, style, hip, micro, rime, rappeurs, phases, lyrics, textes, rappeur baby, soir, might, chérie, yeah, bébé, nuit, miss, night, soirée, copines, danse, fille, mama, belle	14.91	Ärsenik
amour, cœur, yeux, might, larmes, sentiments, fois, nuit, fille, souvenirs, temps, vie, soir, mal, relation	10.35	Dadju
négro, négrlos, ekip, bitch, LDO, Fuck, nigga, pétasse, Dakar, chatte, game, bite, secte, flow, gang	8.92	Freeze Corleone

Le *topic modeling*

Mots caractérisant le topic	%	Nom (selon l'auteur)	Artiste représentatif
beuh, sacoche, binks, billets, poto, monnaie, Audi, moula, zone, condés, euros, pétard, cli, mets, frappe rêves, ciel, espoir, bonheur, vie, larmes, lumière, souvenirs, jour, destin, âme, vide, monde, silence, soleil peuple, pays, système, justice, liberté, jeunes, violence, médias, familles, politique, État, droits, racisme, peuples, pauvres	19.96	Koba LaD	Gangsta rap
rap, mic, flow, MC, rimes, beat, style, hip, micro, rime, rappeurs, phases, lyrics, textes, rappeur baby, soir, might, chérie, yeah, bébé, nuit, miss, night, soirée, copines, danse, fille, mama, belle	16.37	Dooz Kawa	Rap poétique
amour, cœur, yeux, might, larmes, sentiments, fois, nuit, fille, souvenirs, temps, vie, soir, mal, relation négro, négros, ekip, bitch, LDO, Fuck, nigga, pétasse, Dakar, chatte, game, bite, secte, flow, gang	15.47	Keny Arkana	Rap conscient
	14.91	Ärsenik	Egotrip/Méta-rap
	13.43	JuL	Rap coquin
	10.35	Dadju	Chansons d'amour
	8.92	Freeze Corleone	Rap hardcore/Drill



Le *topic modeling*

Mots caractérisant le topic	%	Nom (selon l'auteur)	Artiste représentatif
beuh, sacoche, binks, billets, poto, monnaie, Audi, moula, zone, condés, euros, pétard, cli, mets, frappe rêves, ciel, espoir, bonheur, vie, larmes, lumière, souvenirs, jour, destin, âme, vide, monde, silence, soleil peuple, pays, système, justice, liberté, jeunes, violence, médias, familles, politique, État, droits, racisme, peuples, pauvres	19.96	Koba LaD	Gangsta rap
rap, mic, flow, MC, rimes, beat, style, hip, micro, rime, rappeurs, phases, lyrics, textes, rappeur baby, soir, might, chérie, yeah, bébé, nuit, miss, night, soirée, copines, danse, fille, mama, belle	16.37	Dooz Kawa	Rap poétique
amour, cœur, yeux, might, larmes, sentiments, fois, nuit, fille, souvenirs, temps, vie, soir, mal, relation négro, négros, ekip, bitch, LDO, Fuck, nigga, pétasse, Dakar, chatte, game, bite, secte, flow, gang	15.47	Keny Arkana	Rap conscient
	14.91	Ärsenik	Egotrip/Méta-rap
	13.43	JuL	Rap coquin
	10.35	Dadju	Chansons d'amour
	8.92	Freeze Corleone	Rap hardcore/Drill

“Blanche”, Maes ft. Booba

<Couplet=2>

Je n'connais pas le banc à part celui de la G.A.V

Drapeau est noir et blanc, je reste devant, je préfère être à l'attaque

Si j'n'ai plus d'limite, je m'arrête où ?

J'pends mon linge sale devant personne, je rachète tout

Je ne demande qu'à t'aimer, te montrer mon Makélélé

Le Versace, le D&G, elle est trop bonne, wallaye, c'est B

Je vais l'emmener à l'étranger, peut-être sur place, j'veais l'échanger

Faire le hlel ou l'étrangler, j'suis qu'en léger, que la CB

J'drage à la Momo Henni, beaucoup de cash, je la séduis

On m'tire dessus, j'entends des bruits, même en réserve, je la détruis

J'viens du 9.2, la quality, tu connais d'jà l'état d'esprit

Elle m'a d'mandé combien ça coûte comme si j'avais regardé l'prix

<refrain>

<outro>

“Alphonse Karr”, Freeze Corleone

[Couplet 2]

Alphonse Karr, j'arrive mal fonscar

J'cons mon paki, j'arrive chinois comme Jackie

S/o Kaki, au micro nous sommes comme vos grands-paters

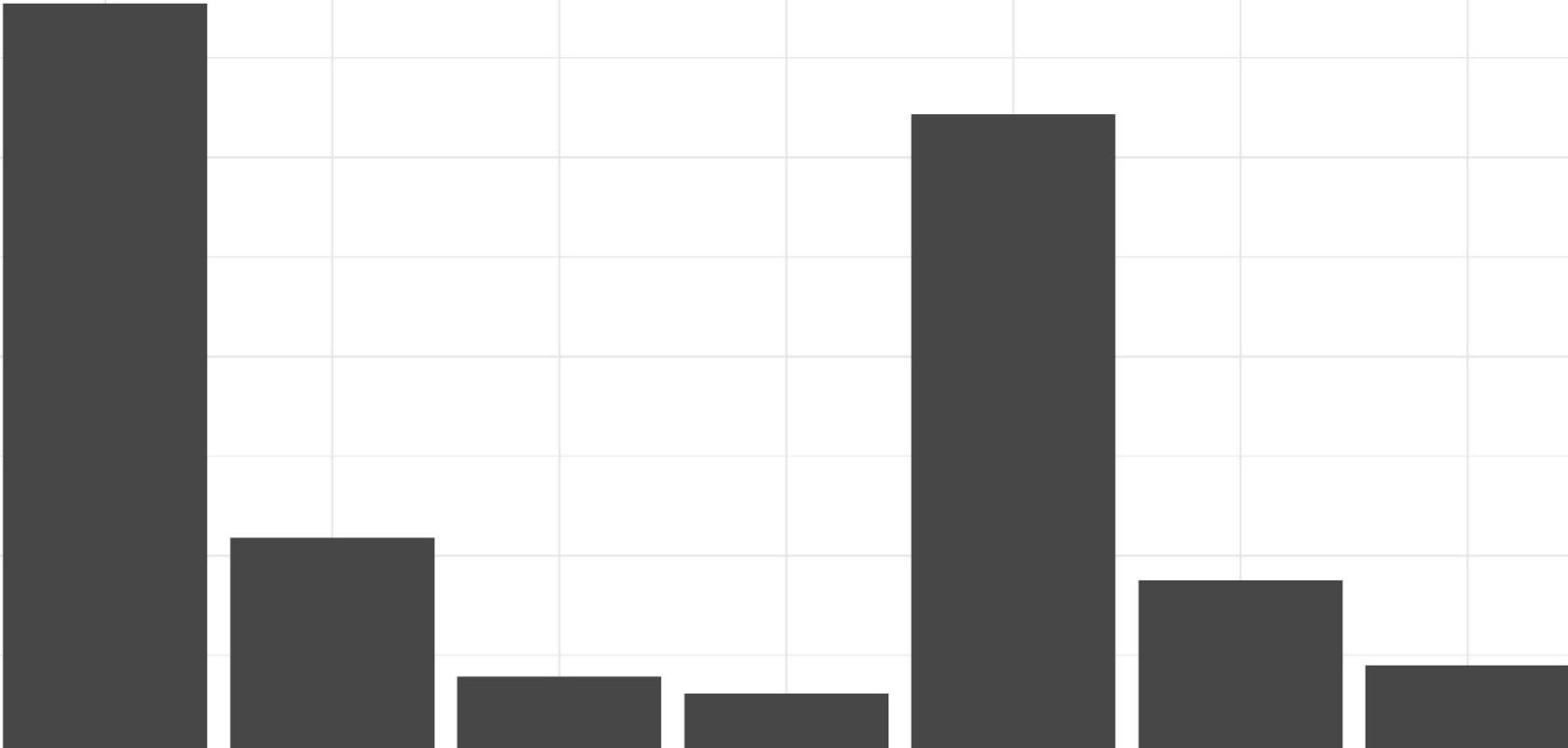
667 ekip des comme nous t'en trouves pas d'autres dans la terre

Fréquence de bébé

0.06
0.04
0.02
0.00

Chansons d'amour Gangsta rap Méta-rap Rap conscient Rap coquin Rap hardcore Rap poétique

Topic



Fréquence de 'beuh'

0.04

0.03

0.02

0.01

0.00

Chansons d'amour Gangsta rap

Méta-rap

Rap conscient

Rap coquin

Rap hardcore

Rap poétique

Topic



III - Une approche confirmatoire : mesurer les ruptures

- Thématique
- Syntaxique
- Énonciative

Blanche, Maes ft. Booba

<couplet=1>

<locuteur=Maes>

Très loin d'être parfait mais t'inquiète pas, ça ira
Au ciel, on va aller, / j'ves laisser mes biens en bas

Toutes mes années d'galère où j'me disais ca ira
Fais un pas de travers, / une bécane, tu t'en iras

J'ves laissé le parquet, / c'est soit Bercy soit du ferme
J'suis dans le carré, / loin d'Bois-d'Arcy, j'ai mon fer

J'barde dans la ville, / sept cents chevaux / Ferrari
Tu connais l'tarif, / fais pas l'grossiste pour un kil'

J'bicrave à pas d'heure, / j'vend la moula dans le bendo
Douze heures à douze heures, / j'fais partir trois kilos d'aya

Ta go' veut un marmot, / moi, j'veoulais juste la ken
Te-shi m'rend paro, / j'me rappelle plus d'la veille

<refrain>

J'vesser la night, elle va s'en remettre, si tu savais
Blanche ou caramel, j'ai son remède

J'vesser la night, elle va s'en remettre, si tu savais
Blanche ou caramel, j'ai son remède

<locuteur=Booba>

Nous restâmes sur le bateau, jamais il ne chavira
Je devais voir Anne-So', au final, j'ai vu Samira

Passe-moi à travers, jamais tu n'me graviras
Tu n'as pas de valeur, jamais on n'se ralliera

<couplet=2>

Je n'connais pas le banc à part celui de la G.A.V
Drapeau est noir et blanc, je reste devant, je préfère être
à l'attaque

Si j'n'ai plus d'limite, je m'arrête où ?
J'pends mon linge sale devant personne, je rachète tout

Je ne demande qu'à t'aimer, te montrer mon Makéléle
Le Versace, le D&G, elle est trop bonne, wallaye, c'est B

Je vais l'emmener à l'étranger, peut-être sur place, j'ves l'
échanger

Faire le hlel ou l'étrangler, j'suis qu'en léger, que la CB

J'drage à la Momo Henni, **beaucoup de cash**, je la séduis
On m'tire dessus, j'entends des bruits, même en réserve, je la
détruis

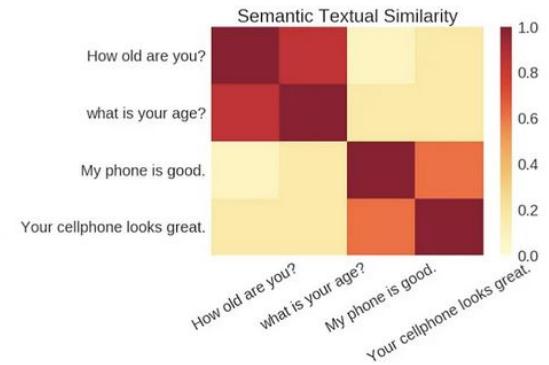
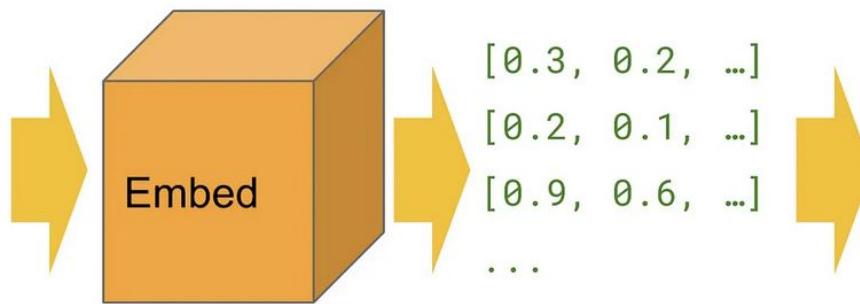
J'vens du 9.2, **la quality**, tu connais d'jà l'état d'esprit
Elle m'a d'mandé combien ça coûte comme si j'avais regardé
l'prix

<refrain>

<outro>

Les “plongements sémantiques”

"How old are you?"
"What is your age?"
"My phone is good."
..."

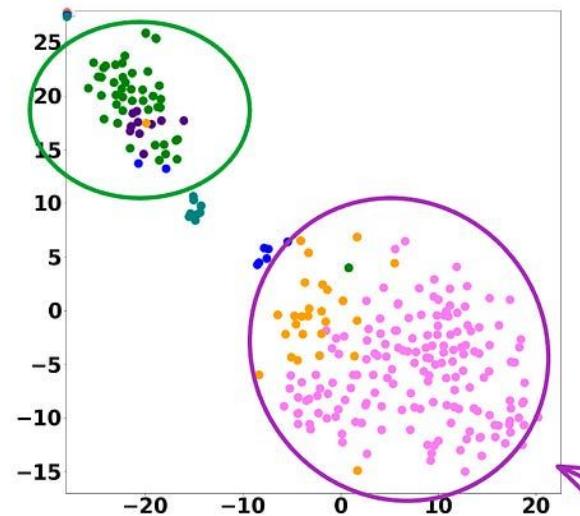


Les “plongements sémantiques”



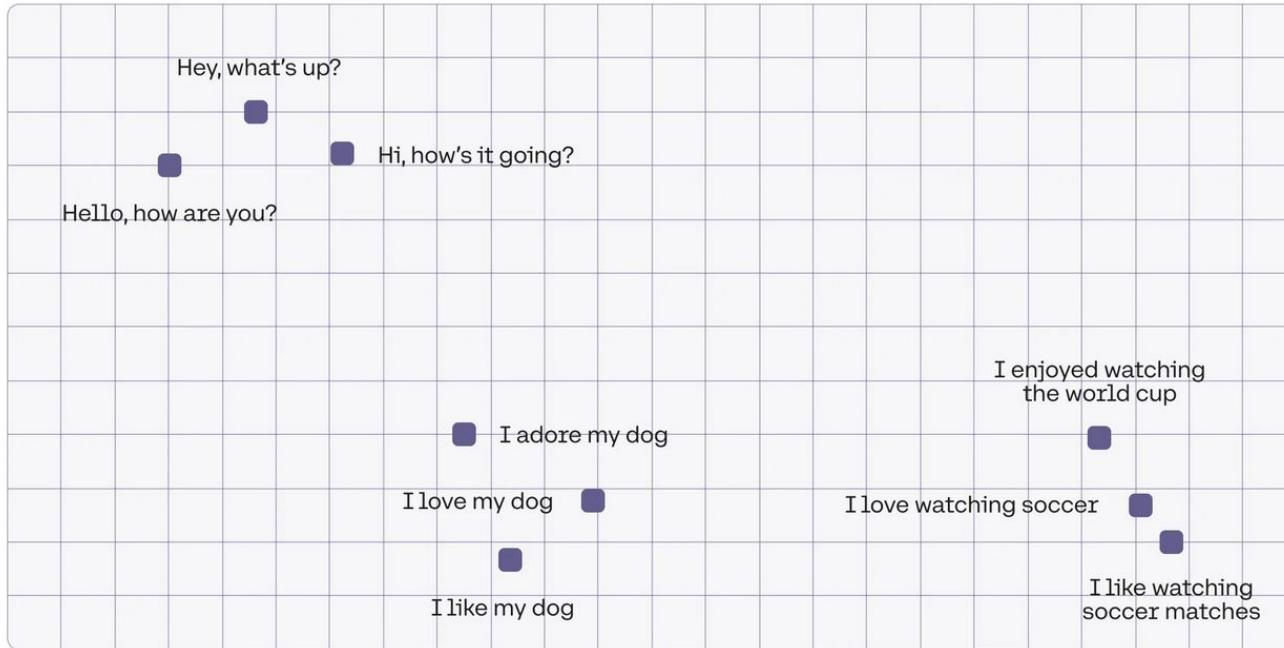
Les “plongements sémantiques”

The word 'bank'
when used in
water sense



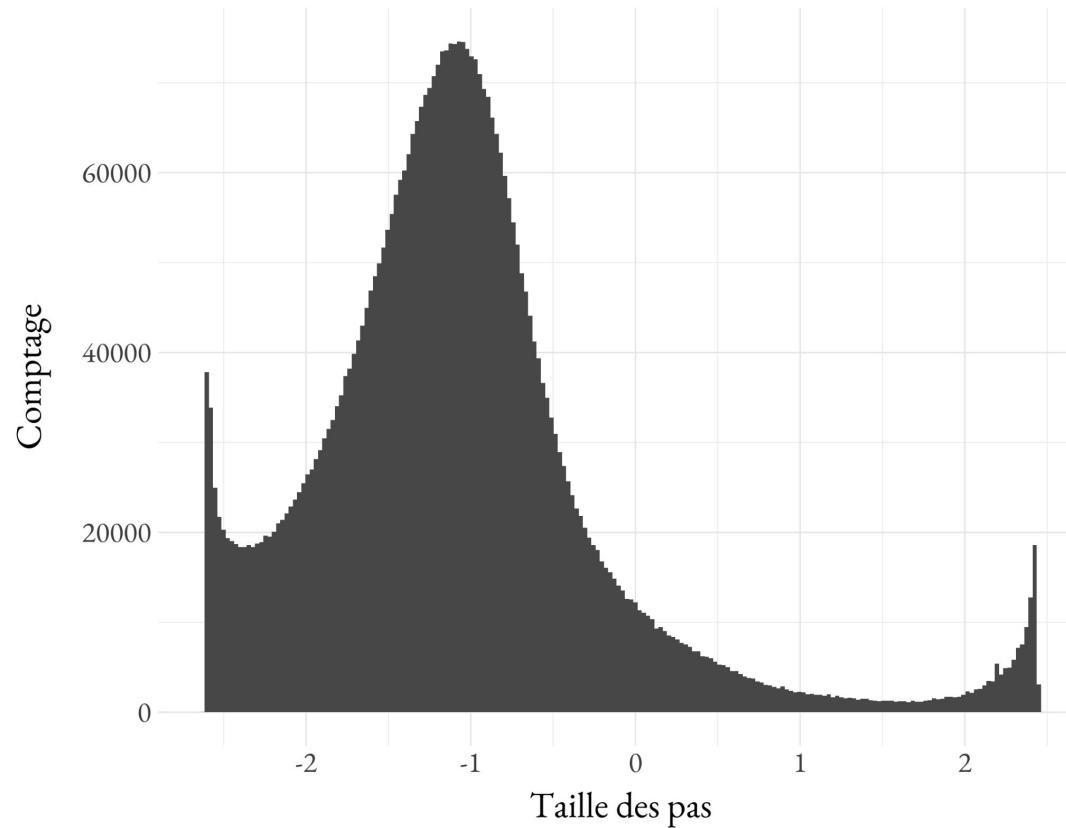
The word 'bank'
when used in
financial sense

Les “plongements sémantiques”

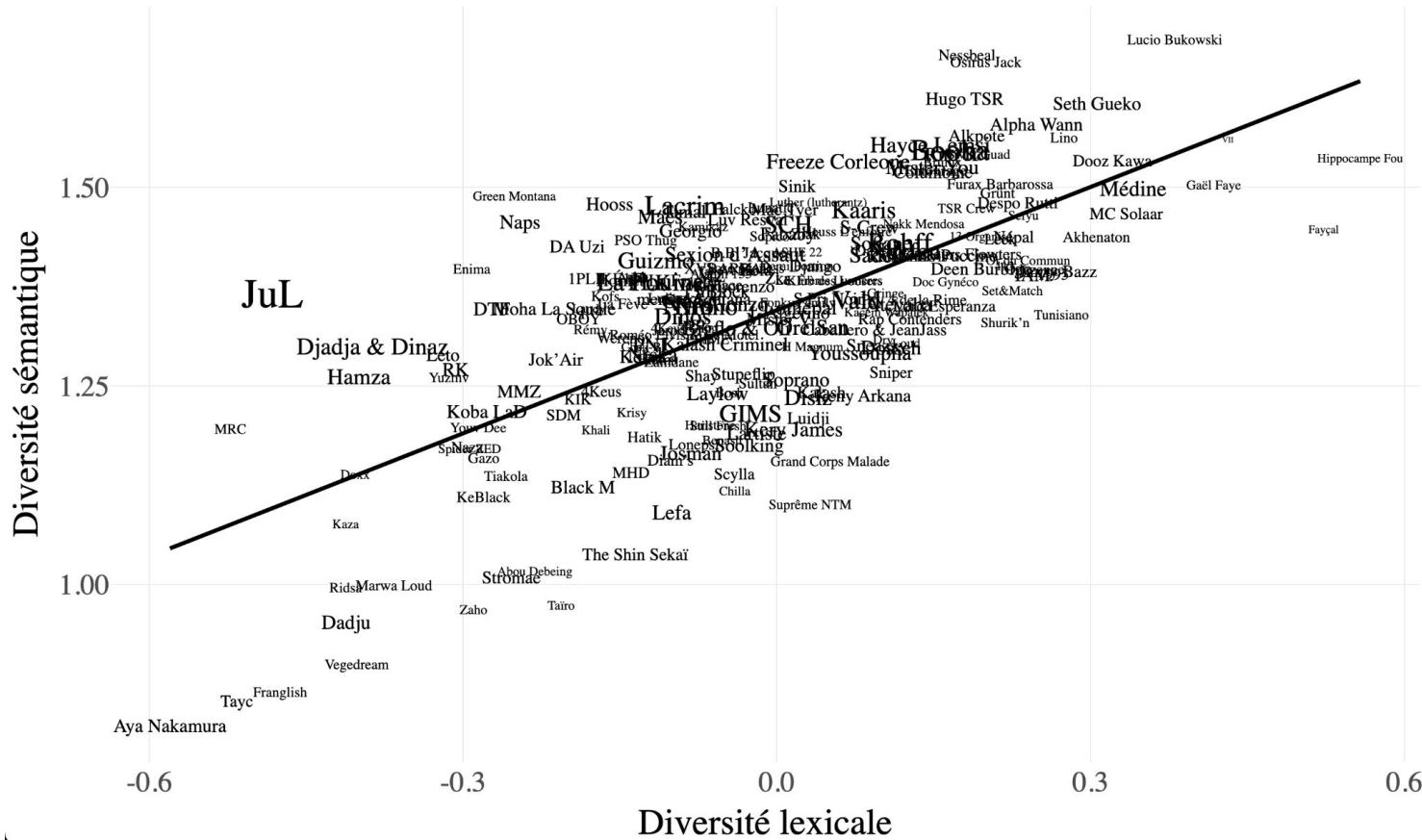


- > On peut mesurer une “distance sémantique” entre deux lignes !
- > Vegedream a la plus faible “taille de pas” moyenne, Lucio Bukowski la plus grande

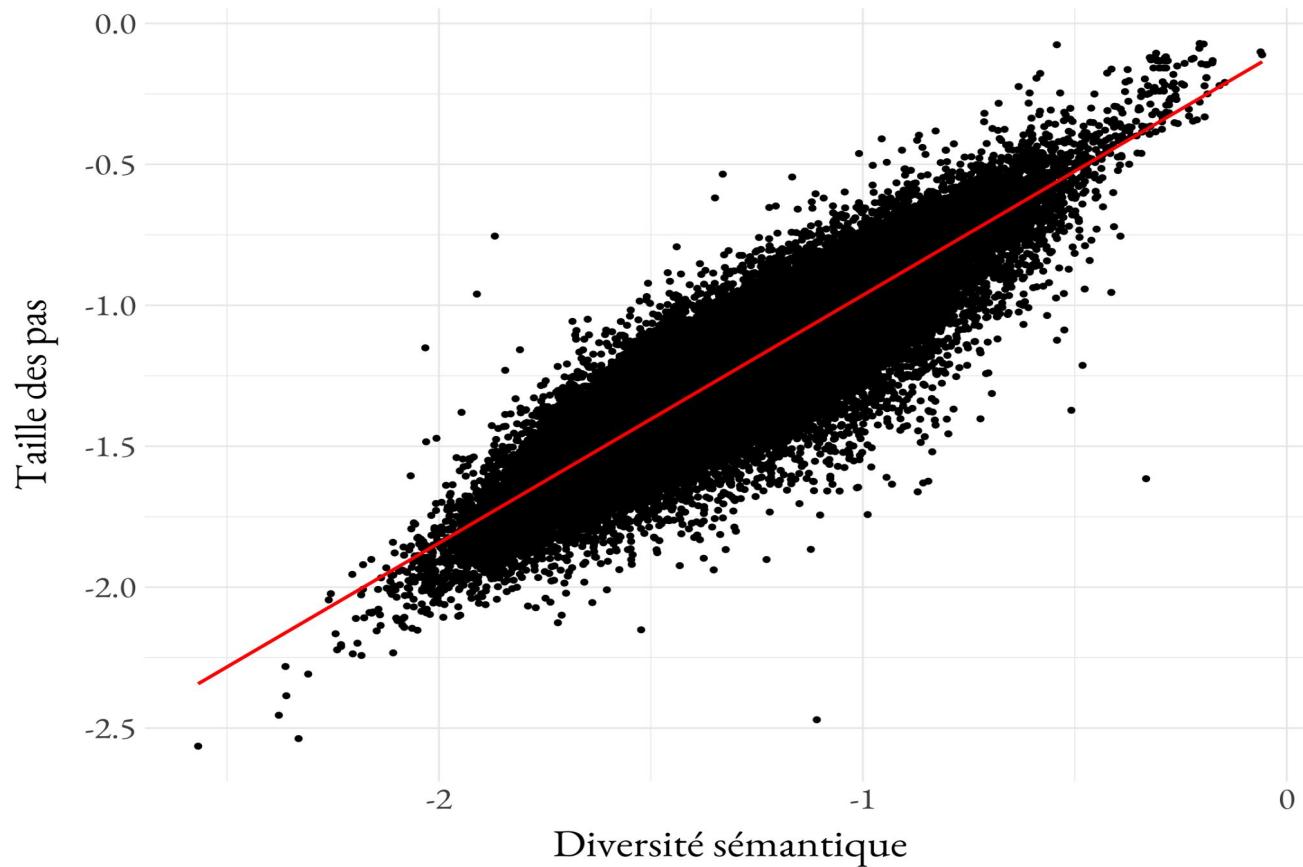
Problème 1 : les refrains



Problème 2 : la diversité sémantique



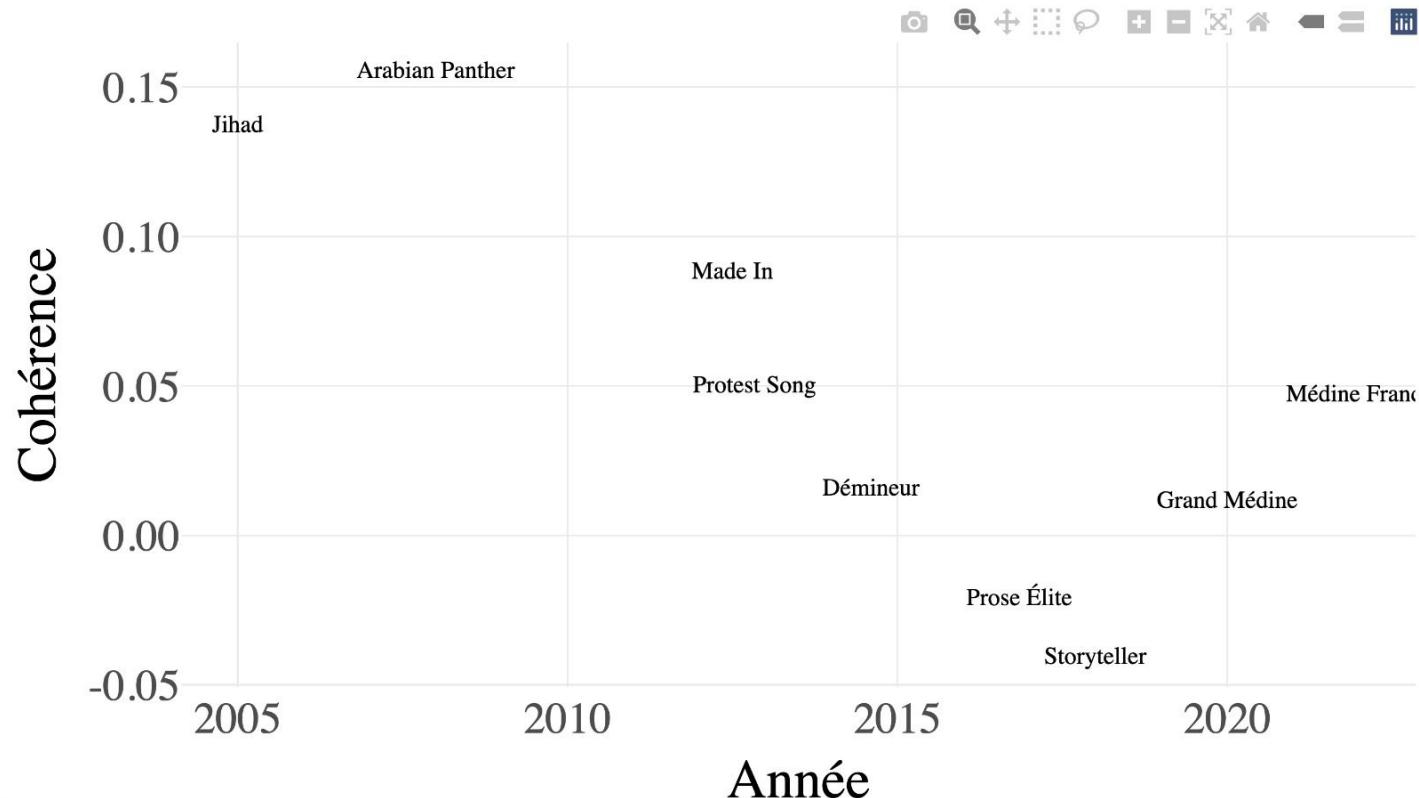
La solution : les “puzzles de mots et de pensée”



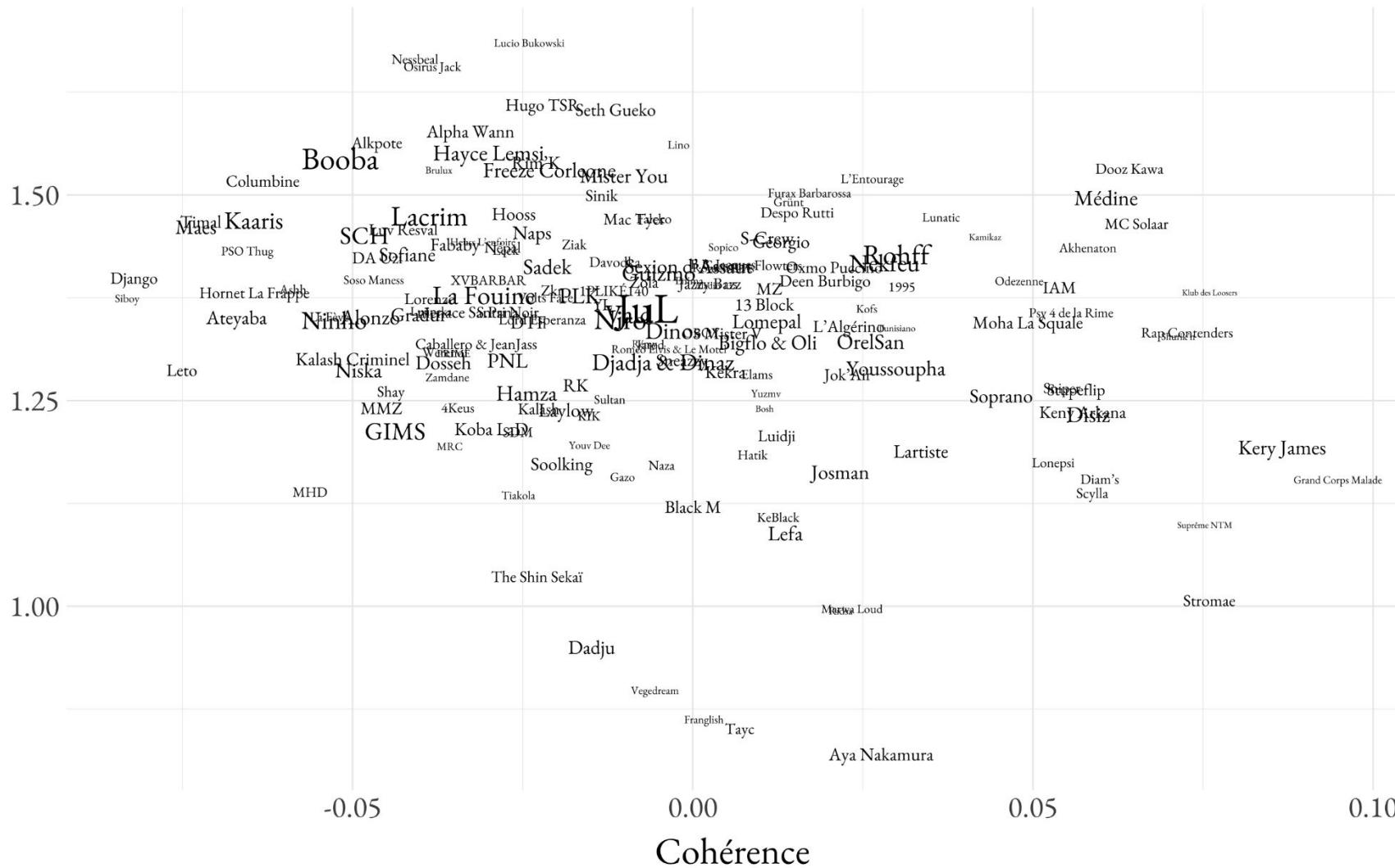
Les chansons les plus cohérentes de Médine

artist	title
Médine	Enfant du destin (Petit Cheval)
Médine	Arabospiritual
Médine	Enfant du destin (Kounta Kinté)
Médine	Sister Act
Médine	Besoin de résolution
Médine	Du Panjshir à Harlem
Médine	Combat de femme
Médine	Enfant du destin (Daoud)
Médine	Biopic

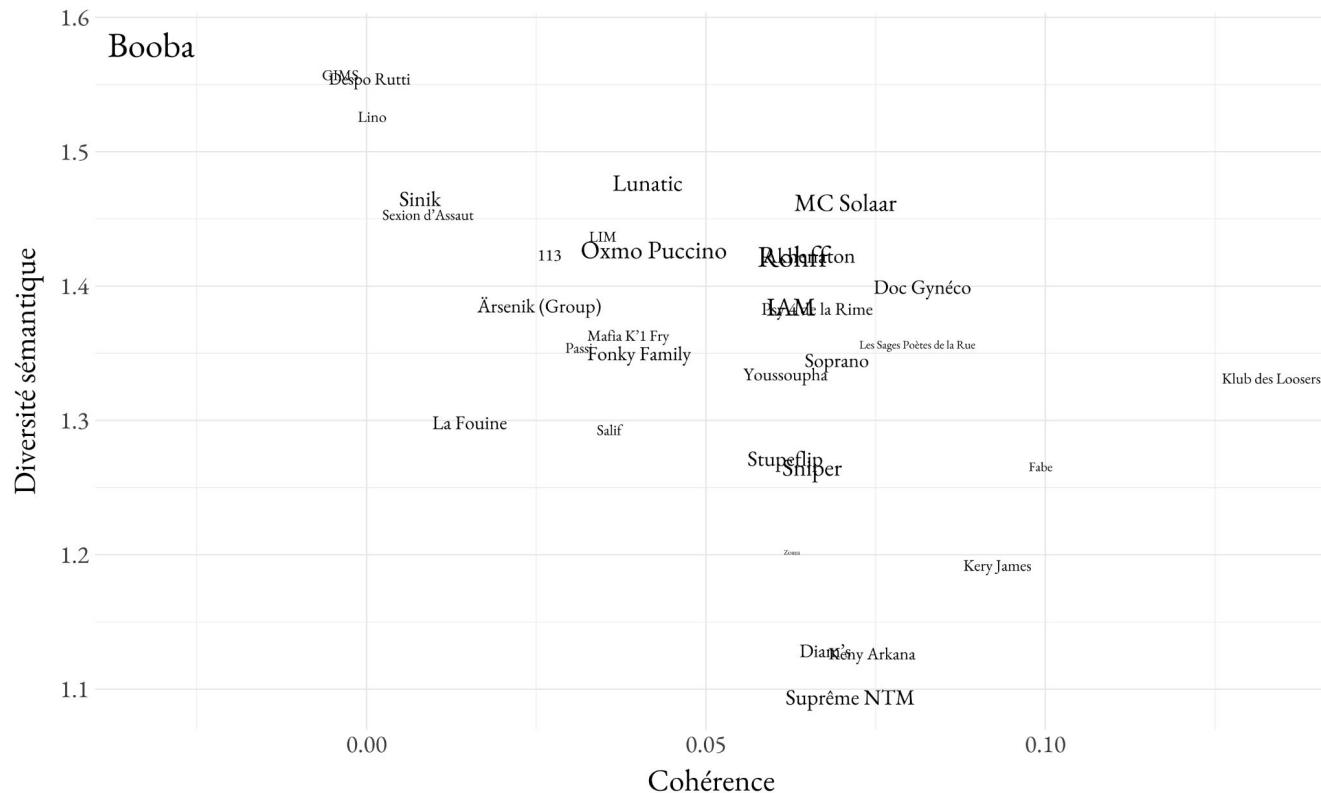
Le virage de la “trap consciente”



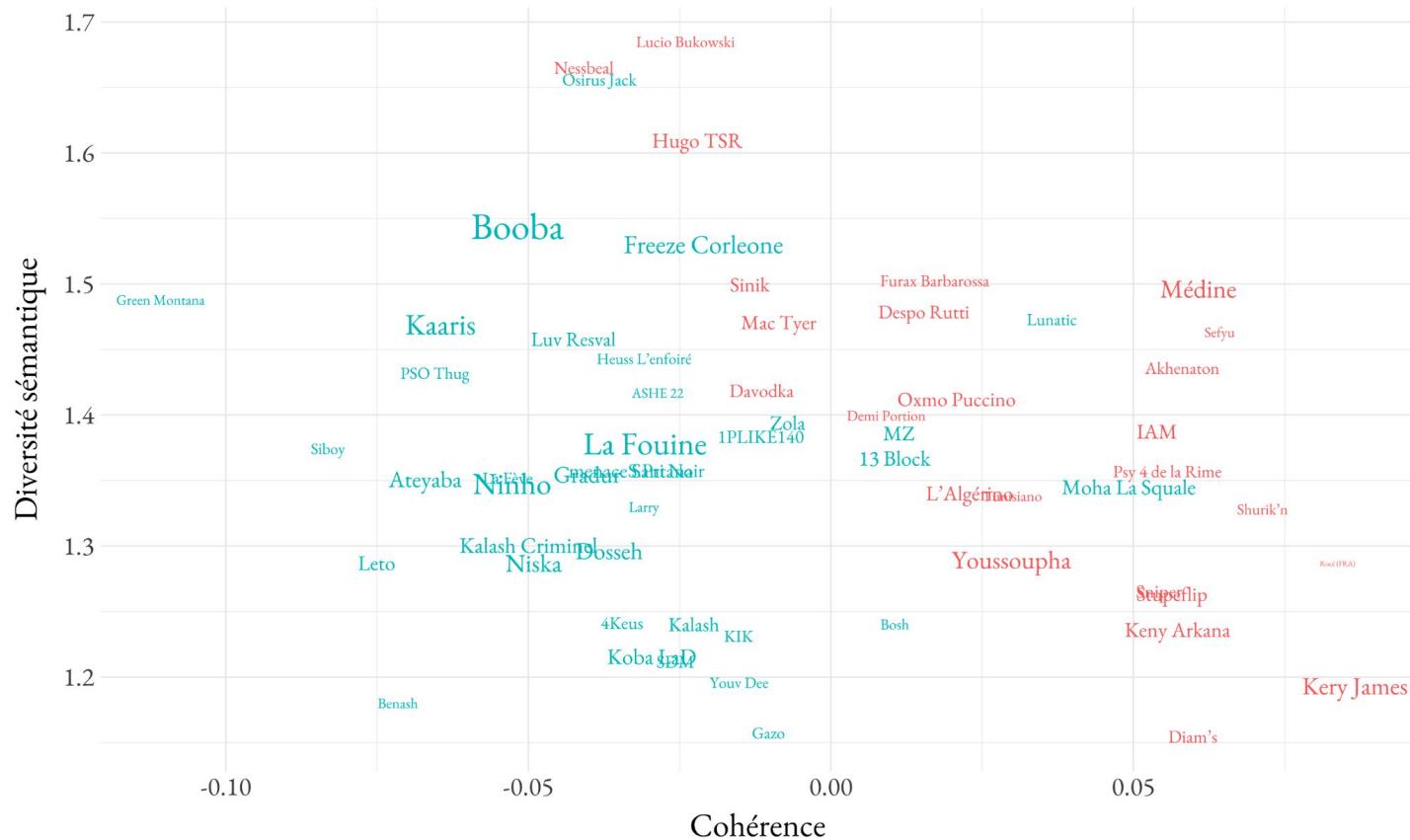
Diversité sémantique



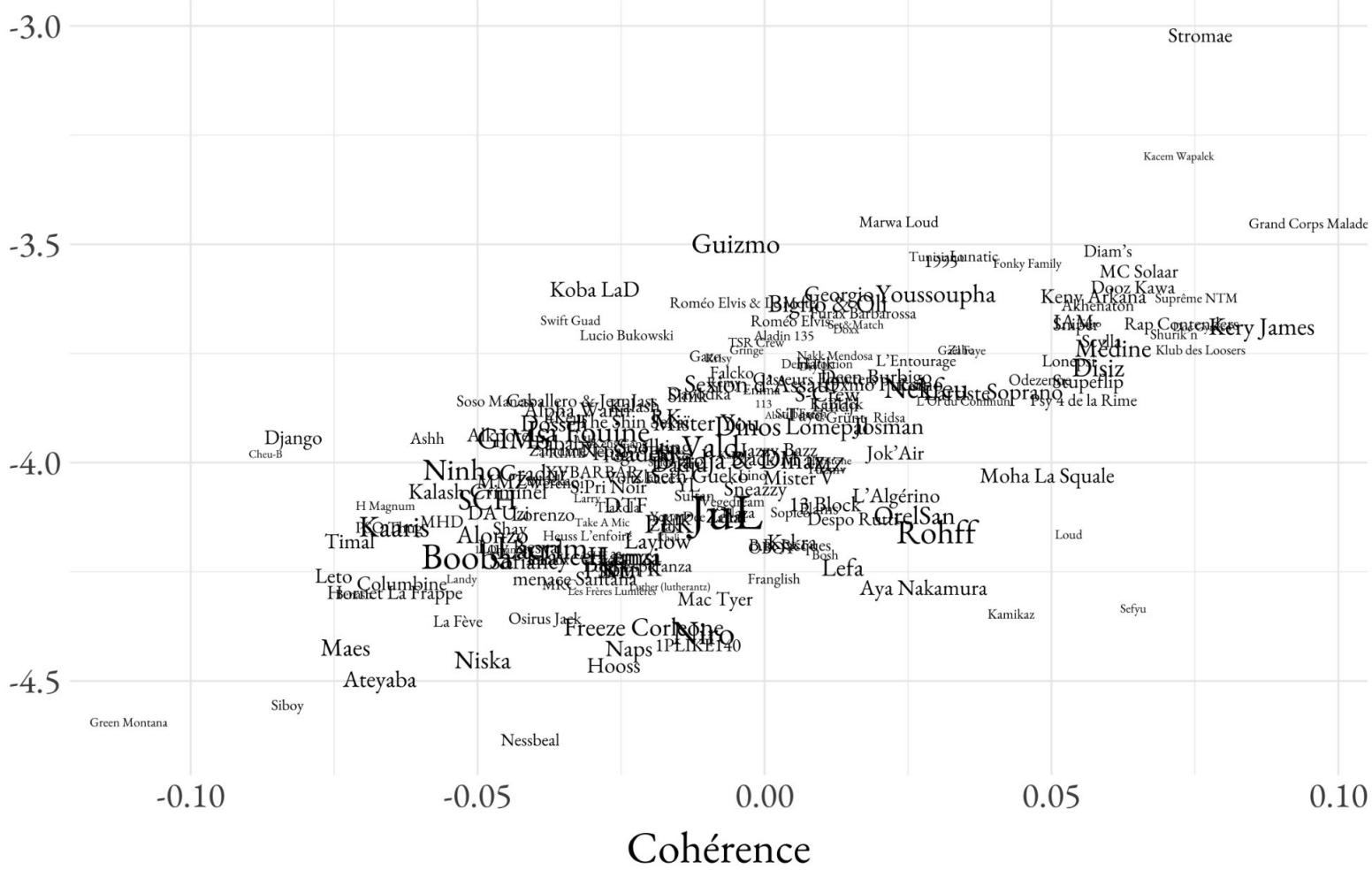
Booba, pionnier de l'incohérence



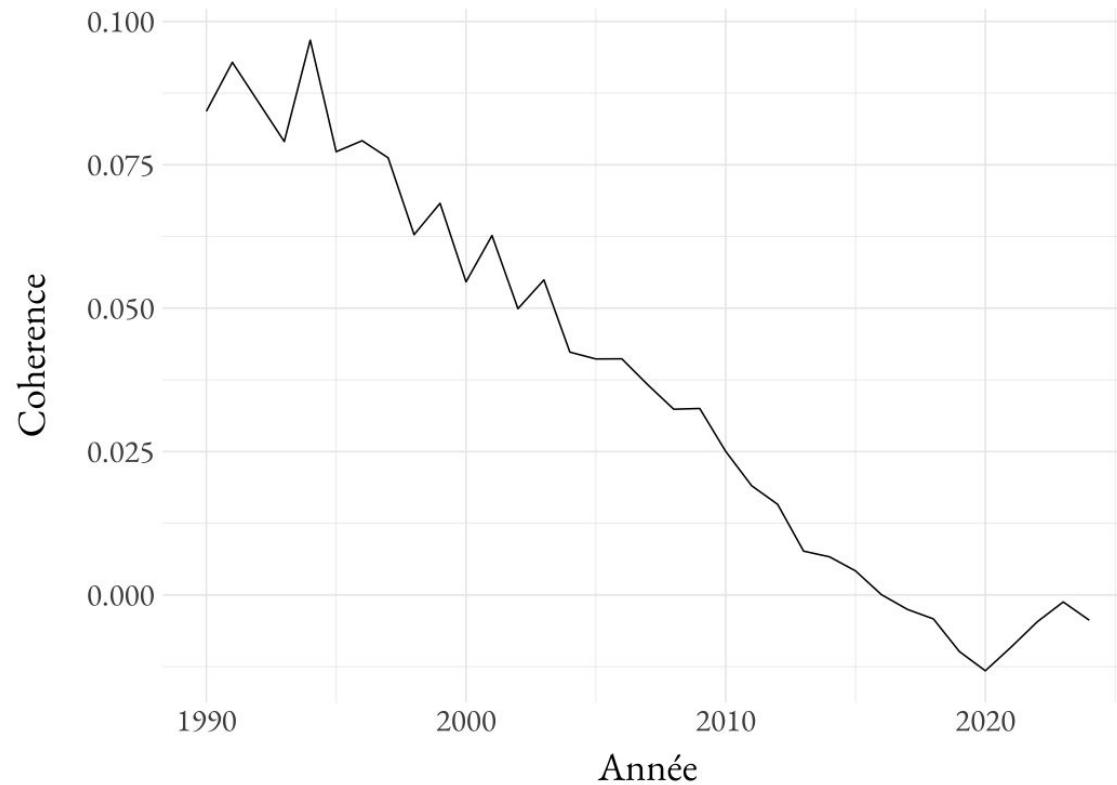
Une dimension structurante



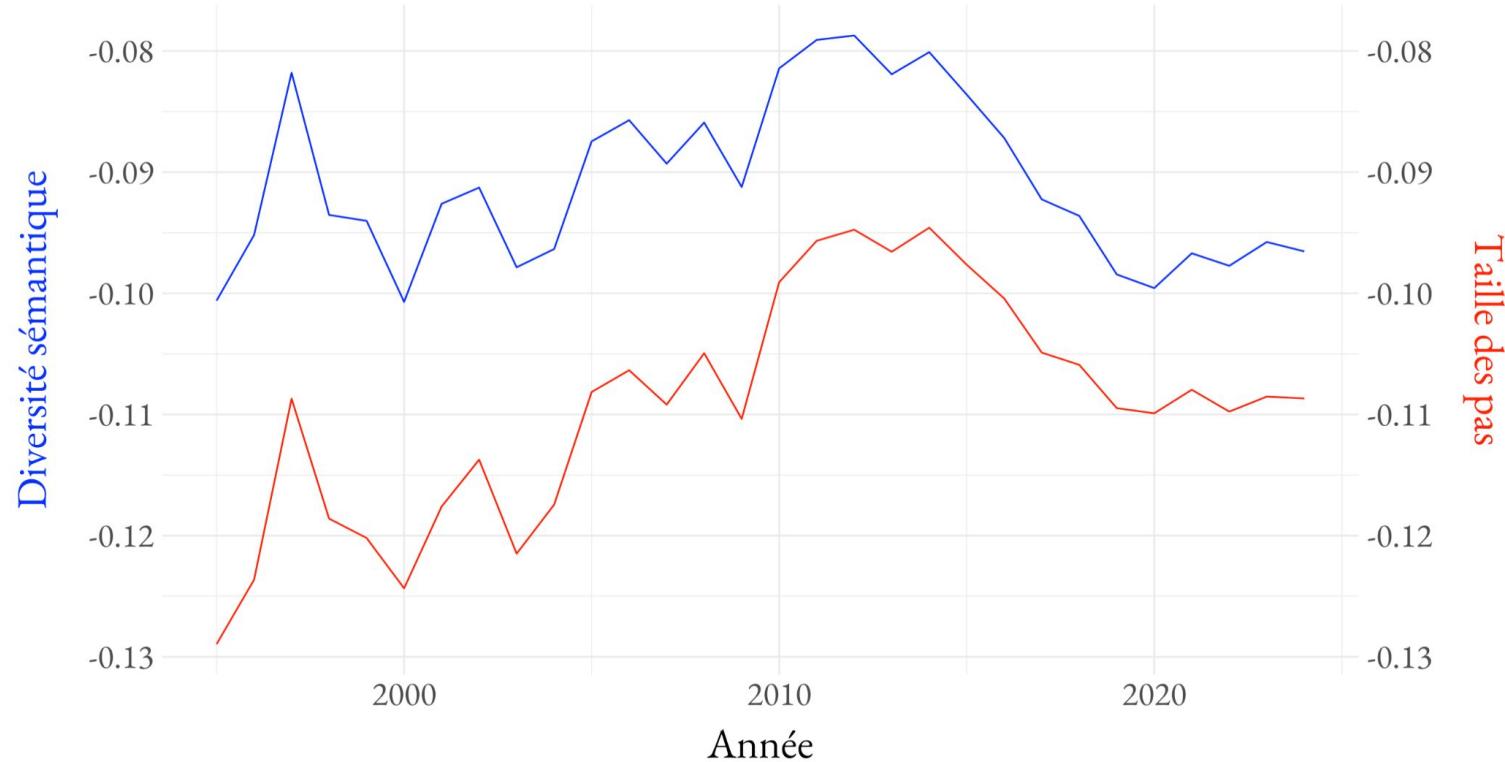
Conjonctions de coordination (log)



Le triomphe de l'incohérence



Le triomphe de l'incohérence



Hardcorité

3

2

1

0

-1

-2

-3

-2

-1

0

1

2

Littérarité

Onomatopées
Interjections

Noms propres

Ateyaba

Haine

Seth Gueko

Alkpote

Apep

Diversité sémantique

Noms
Déterminants
Divinité lexicale
Prépositions

Casey (FRA)

La Rumeur

Odezenne
Dooz
Kawa
Akhenaton
IAM

Keny Arkana

Ali (FRA)

Conjonction de coordination
Anxiété
Mots négatifs

Mots posés
Tristesse
Emotions
Mots français

Lonepsi

The Shin Sekai
Conjonctions de subordination
Adverbes

Stromae

Pronoms

Aya Nakamura

Dadju

Auxiliaires

Black M

Djadja & Dinaz

Lartiste kofis

Bigflo & Oli

Lujipeka

Stromae

Scylla

Scylla

Kery James

Keny Arkana

Casey (FRA)

Osiris Jack

Freeze Corleone

Ateyaba

Alkpote

Rim K

Nessbeal

SPi Noir

XVRBABALZO

Yannick Wann

Zola

Gazo

IPCO

IPCO