

LES ÉDITORIAUX DE MIROIR DU CYCLISME : UNE PREMIÈRE APPROCHE STATISTIQUE

Analyse descriptive et factorielle, modélisation de sujets

Marie Puren, Aurélien Pellet

27 June 2023

Journée IRIS - Projet Collex Persée - MSH de Dijon

- Une première analyse quantitative...
- Enrichie par vos remarques, retours et analyses personnelles.
- Pas d'ambition de résumer l'intégralité du corpus ni de le décrire en détail.
- Les analyses qualitatives et la connaissance approfondie du corpus constituent à la fois le moteur, la source et la "prochaine étape" de nos analyses.

ANALYSE DESCRIPTIVE

PRÉSENTATION DES ÉDITORIAUX

- 372 éditoriaux répartis entre 1960 et 1994.
- Moyenne de 662 mots par document, avec une longueur maximale de 3324 mots.

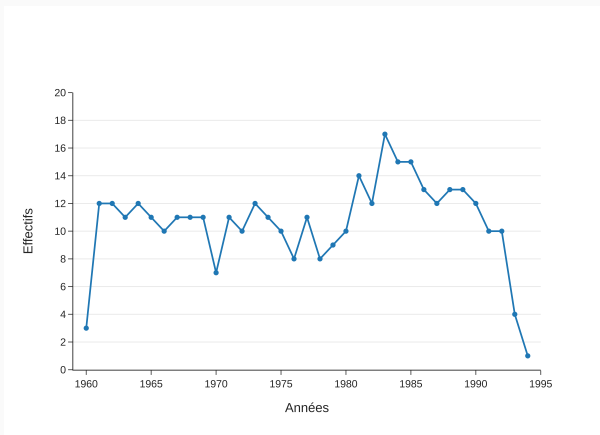


Figure – Répartition des éditoriaux par années

PRÉSENTATION DES ÉDITORIAUX

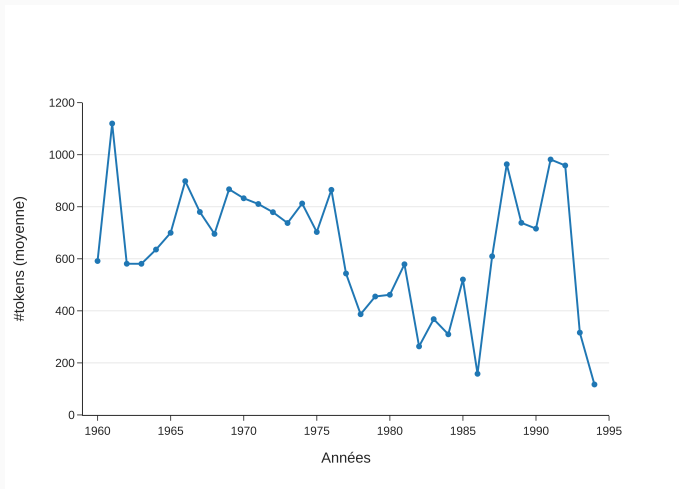


Figure – Nombre moyen de mots par années

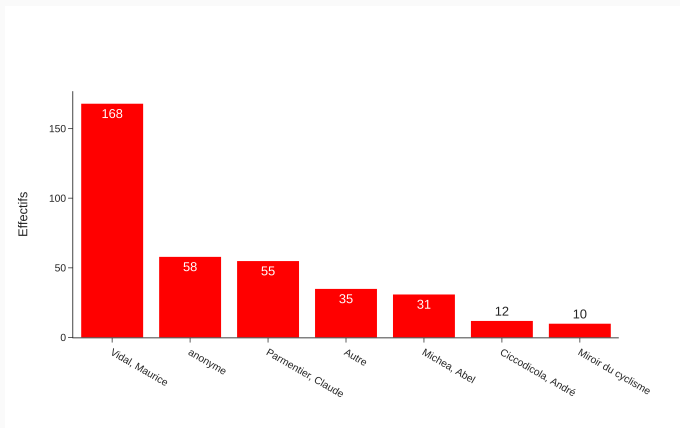


Figure – Répartition des auteurs

LES AUTEURS

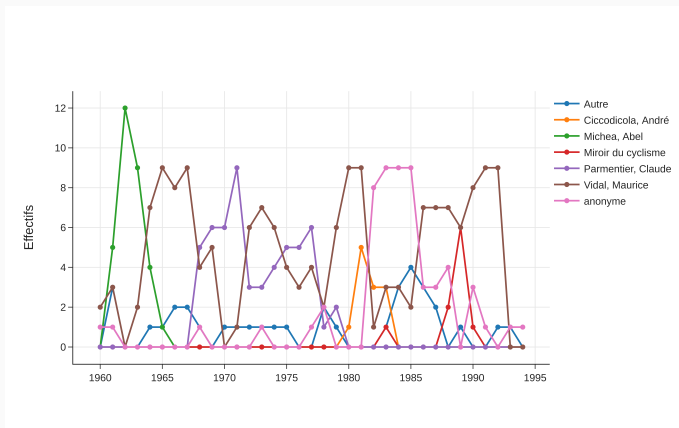


Figure – Les auteurs à travers le temps

Pour mieux étudier notre vocabulaire, nous effectuons quelques étapes de nettoyage :

- Retrait des "stop words" : le, mais, la, quand
- Lemmatisation du vocabulaire
- Fréquence vs Tf-Idf

Les 30 mots les plus fréquents	Les 30 mots les plus "tfidf-fréquents"
'tour', 'cyclisme', 'faire', 'pouvoir', 'france', 'coureur', 'sport', 'grand', 'équipe', 'luire', 'année', 'course', 'cycliste', 'français', 'champion', 'professionnel', 'monde', 'savoir', 'sportif', 'an', 'homme', 'aller', 'voir', 'devoir', 'saison', 'falloir', 'jour', 'épreuve', 'vouloir', 'seul'	'tour', 'cyclisme', 'faire', 'france', 'pouvoir', 'sport', 'coureur', 'grand', 'équipe', 'luire', 'hinault', 'année', 'course', 'saison', 'professionnel', français', 'champion', 'cycliste', 'monde', 'homme', 'bernard', 'miroir', 'an', 'sportif', 'voir', 'savoir', 'devoir', 'aller', 'falloir', 'merckx'

Il y a beaucoup de similarités, mais :

- L'ordre relatif est le plus important.
- Quelques différences : "Hinault", "Miroir" ou même "Merckx".

LES COUREURS

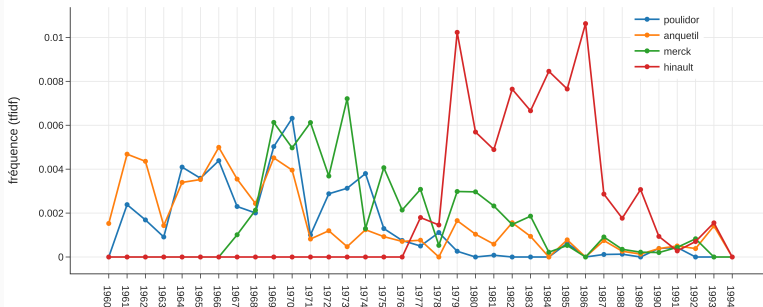


Figure – Poids des mots par années

<u>Corrélations</u>	Poulidor	Anquetil	Hinault	Merck(x)
Poulidor	1	0.51	-0.17	0.04
Anquetil	0.51	1	-0.05	0.03
Hinault	-0.17	-0.05	1	0.02
Merck(x)	0.04	0.03	0.02	1

Perspectives :

- Nous envisageons d'approfondir davantage notre analyse en explorant d'autres méthodes, telles que l'analyse des entités nommées (NER).
- Nous prévoyons également de poursuivre nos recherches sur les co-occurrences et d'examiner plus en détail leurs implications.
- Il sera intéressant d'explorer plus en profondeur le lien entre les auteurs des textes et le type de vocabulaire qu'ils utilisent, afin de mieux comprendre leur style d'écriture.

ANALYSE FACTORIELLE

UN PREMIER EXEMPLE AVEC BOURDIEU

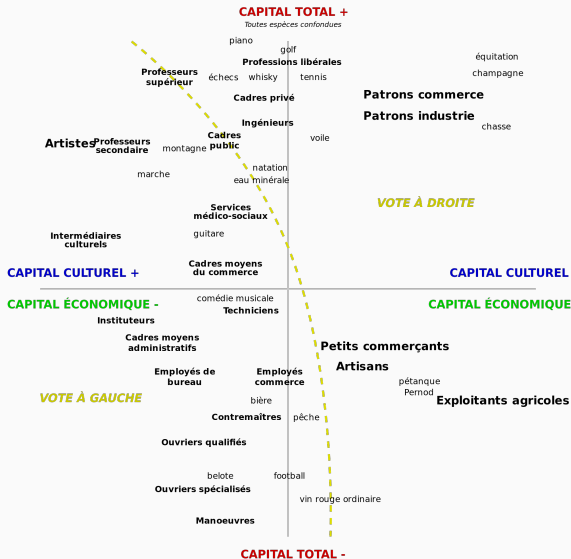


Figure – Analyse en composante multiple de Bourdieu

- L'Analyse en Composantes Principales (ACP ou PCA) utilise la matrice des fréquences (tf-idf).
- L'objectif est de trouver les axes de variance maximale dans nos données, c'est-à-dire les endroits où il y a le plus de variations dans le vocabulaire (zones d'entropie maximale).
- Ces composantes, le long desquelles la variance est maximale, permettent d'identifier des variations significatives dans le vocabulaire et donc de déterminer certaines thématiques.
- Les composantes sont des combinaisons linéaires des mots de notre vocabulaire.

LA PREMIÈRE COMPOSANTE

Mots contribuant le plus à la première composante	Mots contribuant le plus à la seconde composante
'dirigeant', 'groupe', 'organisateur', 'président', 'coureur', 'directeur', 'problème', 'national', 'fédération', 'amateur', 'professionnel'	'loin', 'cyclisme', 'sportif', 'athlète', 'savoir', 'populaire', 'croire', 'bicyclette', 'vie', 'cycliste', 'maurice', 'sport', 'vidal'

Une première thématique intéressante émerge, liée à l'organisation du circuit de courses, aux fédérations, aux décisions et aux dirigeants.

Nous pouvons identifier des documents fortement corrélés à cet axe :

- L'éditorial n°184 d'avril 1974, qui évoque un conflit auquel la FICP a dû faire face.
- L'éditorial n°311 de 1981, qui discute de l'élection d'un nouveau président à la tête de l'UCI.

La question de l'organisation du cyclisme est une thématique centrale dans les éditoriaux, et il ressort souvent un conflit entre le circuit professionnel et la pratique amateur, comme l'évoque l'éditorial dont nous venons de parler.

Les mots négativement corrélés à cette composante sont : cyclisme, sportif, populaire, bicyclette, sport.

Cette composante semble être liée au traitement du cyclisme du point de vue de l'amateur et/ou du spectateur. On parle de la pratique du cyclisme, de sa popularité, des mesures prises par l'État pour favoriser l'utilisation du vélo, ainsi que du regard du spectateur sur le monde du cyclisme. Cette composante semble s'intéresser à ceux qui regardent ou pratiquent le vélo, mais avec un point de vue extérieur au monde professionnel.

Voici quelques exemples :

- L'éditorial n°58 de 1965 aborde la situation du cyclisme en France et son écart par rapport à d'autres sports populaires.
- L'éditorial n°410 parle, quant à lui, du regard d'un spectateur sur le Tour de France, des difficultés de percevoir ou d'approcher les coureurs, qui sont tous très encadrés.

DEUX NOUVEAUX EXEMPLES

l'avenir.

Ce désintéressement pour le cyclisme professionnel qui frappe les constructeurs et les extra-sportifs trouve son explication dans plusieurs raisons. Tout d'abord, la conjoncture économique se prête assez peu aux investissements et il est certain que depuis des lustres, ceux qui ont contribué à l'essor, ou plus tard à la survie des groupes n'ont jamais su tirer profit de l'argument publicitaire dont ils disposaient. Il n'est que de voir leur publicité ou de constater dans quel ignorance se trouve le grand public quand on l'interroge sur les raisons sociales d'une firme extra-sportive qui équipe tel ou tel coureur. A n'avoir pas su « faire mousser » leur propre marchandise, ils en viennent tout à la fois à condamner le sport cycliste en général et les coureurs, ces « trop-payés », en particulier. Pour avoir voulu suivre une politique de vedettariat, ils se retrouvent aujourd'hui, eux aussi, à la merci des résultats d'un seul porte-drapeau, responsable de la bonne marche de toute l'entreprise. On pense ainsi à quelques clubs de football, qui négligent la formation des jeunes et vont recruter leurs vedettes à l'extérieur pour obtenir des résultats immédiats à des prix défiant toute concurrence en gravant des budgets déjà hypothéqués par des abus de toutes sortes. La notion du club s'est dégradée au fil des années, celle des équipes ne subsiste plus que par des notions financières discutables, et le Tour de France disputé par équipes nationales n'existe plus pour rappeler aux coureurs que le cyclisme est un sport qui n'est pas toujours soumis totalement à la commercialisation.

(a)

Nul ne peut contester l'immense regain de popularité de la bicyclette, attesté par les chiffres de vente de toutes les firmes françaises et étrangères. A quoi est-il dû ? Sans doute au fait que la « petite reine » a terminé sa maladie infantile : elle avait conquis les foules comme moyen de transport d'une géniale simplicité, d'une rare perfection et d'un coût particulièrement modique. L'automobile et plus encore peut-être le vélomoteur semblaient devoir mettre son existence en danger. Et il est vrai que, les folies de la circulation venant s'ajouter à la motorisation, on vit disparaître de nos cités les innombrables cohortes de cyclistes des petits matins ouvrables. Dans les pays industrialisés comme les nôtres, il est peu probable que le vélo retrouve cette vocation utilitaire (encore que la crise éventuelle du pétrole...) qu'il conserve dans de nombreux et immenses pays en voie de développement. Mais les excès même de cette industrialisation amènent des millions de femmes et d'hommes à découvrir et redécouvrir quel merveilleux moyen de loisir et de santé il constitue. Il suffit de regarder autour de soi pour compter chaque jour de nouveaux adeptes de l'exercice et de la promenade à bicyclette. Le phénomène n'est pas propre à la France, puisqu'en 1972 on a vendu aux Etats-Unis plus de bicyclettes (13 millions) que de voitures (11 millions 1/2). Il faut encore ajouter que le succès est si foudroyant que les néophytes deviennent très vite des amateurs éclairés... et exigeants : à preuve le développement parallèle et considérable de la vente de bicyclette de course équipées des derniers perfectionnements de la technique. Et le snobisme n'explique pas tout... A cette formidable poussée des amateurs de vélo (qui vont vraiment de 7 à 77 ans), comment répondent les pouvoirs publics ? Pour l'instant, par le mépris le plus total. Ne parlons pas des rues de Paris ou des grandes villes où les automobilistes névrosés (pas par leur faute il est vrai) considèrent les « deux roues » comme un gibier suicidaire. Mais les lieux traditionnels de promenade tendent les uns après les autres à devenir des autodromes interdits aux

(b)

Figure – Deux documents avec une forte contribution (resp.) pour les deux premiers axes

Mots contribuant le plus à la première composante	Mots contribuant le plus à la seconde composante
Miroir, tour, numéro, luire	Tour, formule, france, organisateur, jeune, italie, giro, étape

Table – Contribution > 0.01

Deux nouvelles thématiques se dégagent :

La première est liée aux éditoriaux qui parlent d'eux-mêmes. Ceux qui évoquent des changements dans la formule, des nouveautés. Mais aussi des éditoriaux plus courts qui se contentent de décrire les sujets du numéro. Il s'agit de textes qui s'éloignent des analyses et des débats d'actualité.

La seconde composante concerne les grandes courses, telles que le Tour de France et le Giro. On y retrouve des bilans, des analyses et des débats qui entourent ces événements majeurs du cyclisme.

Ces deux thématiques offrent de nouvelles perspectives pour approfondir l'étude de notre corpus d'éditoriaux.



LES COMPOSANTES 1 ET 5

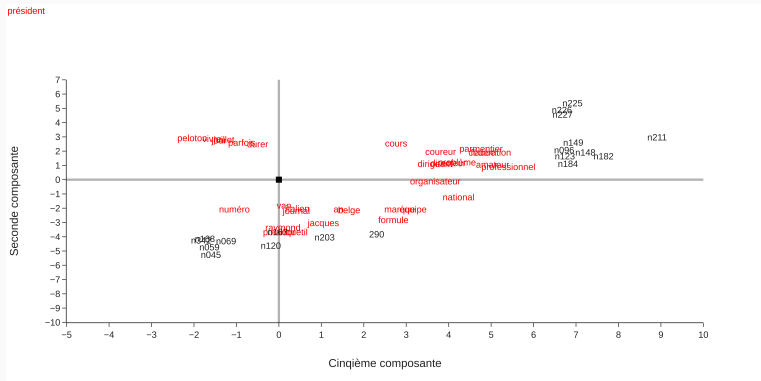


Figure – Composantes principales 1 et 5

Mots contribuant le plus à la première composante	Mots contribuant le plus à la cinquième composante
'dirigeant', 'groupe', 'organisateur', 'président', 'coureur', 'directeur', 'problème', 'national', 'fédération', 'amateur', 'professionnel'	'Poulidor', 'Anquetil', 'Raymond', 'Jacques'

UN DERNIER AXE INTÉRESSANT : POULIDOR-ANQUETIL

Nous observons ici un axe d'analyse qui tourne autour de la rivalité entre Anquetil et Poulidor. L'analyse des fréquences nous avait déjà montré une forte corrélation entre ces deux coureurs. L'analyse en composantes principales (ACP) confirme que cette rivalité est un sujet caractéristique abordé dans certains éditoriaux.

Dans la distribution des rôles, Poulidor n'a pas le meilleur. Sans cesse on lui demande d'attaquer Anquetil, et lui seul sait ce qu'il en coûte. Pourtant, avec cette fois une générosité sans mélange, il a répondu à l'attente. Et s'il a une nouvelle fois été battu, c'est que son vainqueur est par malchance (pour lui) le coureur le plus étonnant des vingt dernières années. Et si je dis que Poulidor s'est encore rapproché de Jacques Anquetil, c'est dire que je le tiens lui-même pour un très grand champion.

(a) Éditorial n°59

Raymond Poulidor admettait la supériorité d'Anquetil et c'était pour lui son meilleur doping. Cette rivalité, soigneusement entretenue pour les besoins de la cause, l'obligeait à se surpasser, à aller au-delà de ses possibilités.

Aujourd'hui, il faut craindre que le départ d'Anquetil n'ait provoqué une cassure chez Raymond Poulidor. Lui-même en est conscient même s'il préfère le cacher.

Poulidor est-il condamné à être toujours le second de quelqu'un ? C'est une question que l'on est en droit de se poser.

Aujourd'hui, c'est Van Springel qui était sur sa route, demain ce sera Merckx, Pingeon ou quelqu'un d'autre. A moins que...

(b) Éditorial n°120

Figure – Deux documents avec une forte contribution sur l'axe 5

Une analyse des cinq axes qui contribuent le plus à l'ACP nous a révélé cinq thématiques principales. Ces thématiques sont considérées comme importantes par l'ACP, car elles expliquent le plus de variance dans nos données.

Ces thématiques comprennent :

- Une thématique liée à l'organisation de la scène professionnelle du cyclisme.
- Une thématique liée à la pratique du cyclisme et au rôle du spectateur.
- Une thématique en lien avec les grandes courses et les événements majeurs du cyclisme.
- Une thématique correspondant à des éditoriaux plus courts qui se contentent de décrire les sujets du numéro.
- Une thématique concentrée sur la rivalité entre Poulidor et Anquetil.

Il est également possible d'analyser d'autres composantes, bien que leur contribution à la variance de nos données soit moins significative.

Perspectives : Une perspective intéressante serait d'explorer la relation entre ces thématiques/axes et les auteurs des éditoriaux, ainsi que de les relier à la période d'écriture. Cela permettrait d'approfondir notre compréhension des différentes perspectives et des évolutions dans les éditoriaux au fil du temps.

TOPIC MODELLING : LDA

-> LDA : Latent Dirichlet Allocation

- La méthode LDA permet de représenter le corpus en termes de sujets (distribution sur les mots).
- Chaque document est représenté par une distribution de sujets, conditionnellement aux mots observés.
- Le nombre de sujets doit être déterminé à l'avance.

Label	Topics	Examples of words	Weight in the corpus
Names of MPs	0,5,10,14,18, 23,35,37,39	Duval, Sigismond, Jules, Martin	0.101
government/parliament	1,6,9,13,17, 19,22,36,38, 41,45,46,49	tribune, projet, adoption, majorité	0.284
economy	2,4,16	agriculture, commerce, patente, betterave	0.069
working class	7,8,31,34	travailler, salaire, usine, mutuelle	0.070
army	11,48	général, régiment, contrôle, militaire	0.041
department	12	Calais, Alpes, Saône, Charente	0.006
trains/communications	15,44	télégraphe, ingénieur, train, travaux	0.069
local politics	20, 33	ville, arrondissement, local, département	0.030
law enforcement	21,40	police, préfet, tribunal, délit	0.055
school	24	lycée, faculté, classe, enfant	0.023
alcohol	25	bouilleur, degré, raisin, octroi	0.019
budget	26,29,30,43	chiffre, budget, dépense, exercice	0.097
colonies	28	métropole, juif, algérien, tonkin	0.018
navy	32	marin, flotte, mer, bâtiment	0.021
building works	27, 42	construction, théâtre, hectare, terrain	0.024
foreign affairs	47	puissance, Madagascar, Angleterre, traité	0.034

Figure – Exemple de sujets obtenus pour le projet AGODA

-> Étant donné que le corpus est relativement petit, l'ajout de données supplémentaires pourrait être bénéfique... et nous en avons!

Phase de nettoyage

- Suppression des mots vides (stop words).
- Lemmatisation.
- Découpage en paragraphes plus petits (par numéros de pages et sauts de lignes).
- Conservation uniquement des caractères alphanumériques et des documents contenant plus de 100 caractères.

Mots les plus fréquents dans tous le corpus (réduit+nettoyé)

'victoire', 'dernier', 'devoir', 'gagner', 'jean', 'aller', 'année',
'point', 'jour', 'naître', 'champion', 'monde', 'bel', 'équipe',
'cyclisme', 'an', 'coureur', 'luire', 'course', 'paris', 'van', 'grand',
'prix', 'france', 'étape', 'pouvoir', 'faire', 'kilomètre', 'heure',
'tour'

LDA SUR TOUS LE CORPUS

Index	15	16	17	18	19	topic_mean
Topic28	coureur	equipe	pouvoir	tour	faire	0.434538
Topic10	velo	pouvoir	cyclisme	sport	cycliste	0.117156
Topic23	luire	velo	pouvoir	aller	faire	0.0899594
Topic1	laurent	pignon	luire	faire	tour	0.0829597
Topic24	paris	victoire	eddy	merck	tour	0.0774483
Topic21	grand	france	luire	anquetil	tour	0.0576734
Topic8	pouvoir	faire	etape	hinault	tour	0.0293387
Topic20	devoir	coureur	piste	faire	pouvoir	0.0287928
Topic15	rik	pouvoir	belge	looy	van	0.0194288
Topic11	paris	adresse	numero	cyclisme	nitroir	0.009521
Topic7	anquetil	piste	heure	record	kilometre	0.00700154
Topic27	col	juillet	heure	etape	kilometre	0.00693793
Topic14	chef	jean	tel	paris	nitroir	0.00570601
Topic9	france	championnat	cyclo	monde	cross	0.00564104
Topic22	critériun	paris	etape	tour	prix	0.0046774
Topic19	maillot	eric	bernard	kilometre	jean	0.00425804
Topic0	france	etape	anquetil	tour	van	0.00308997
Topic3	jean	numero	nitroir	cyclisme	rue	0.00293908
Topic13	permettre	roue	selle	cadre	velo	0.0021497
Topic25	merck	prix	van	kilometre	heure	0.00177635
Topic16	classement	belg	fr	kilometre	point	0.00168589
Topic18	notta	an	giondi	kilometre	it	0.00159239
Topic12	che	bel	vc	tour	pro	0.00149847
Topic4	classement	belg	van	kilometre	heure	0.00100179
Topic26	urs	manche	battre	finale	cycle	0.00029641
Topic5	nl	neo	pro	ex	naître	0.000725325
Topic2	colnago	taille	total	marque	prix	0.000543073
Topic17	narselle	bordeaux	st	paris	vict	0.00049543
Topic6	campagnolo	shinano	pédaller	dérailleur	roue	0.000328115
Topic29	esp	ita	kn	bel	fra	0.00025796

- Nous disposons maintenant d'un peu plus de 19 000 documents sur lesquels nous entraînons le modèle LDA.
- Dans l'ensemble, les topics obtenus sont décevant mais...
- Les 6 topics les plus fréquents dans nos éditoriaux ne sont pas très significatifs.
- Il serait possible d'améliorer les résultats de LDA en modifiant le corpus ou en regardant plus de mots par topics
- Cependant, nous avons décidé de mettre un terme à cette première partie de l'analyse.

Figure – Topics obtenus

+ **Word2Vec** : On associe un vecteur à chaque mot de notre corpus, les mots dans un contexte similaire ont des vecteurs similaires.

+ Cette méthode permet de résoudre des analogies :

- "Femme est à Reine ce que Homme est à ... Roi
- "Paris est à France ce que Tokyo est au ... Japon

L'apprentissage des vecteurs est fortement dépendant du corpus

+ **Doc2Vec** : Algorithme similaire, mais on rajoute un token

"document" afin d'apprendre un vecteur sur chacun des documents

+ **Top2Vec** : Combinaison des deux précédentes méthodes pour identifier des sujets

+ **Limites** : Prise en compte du contexte limitée (15 mots), un mot équivaut à un vecteur, les analogies peuvent être moins efficaces sur de petits corpus

TOPIC MODELLING : PROJECTION

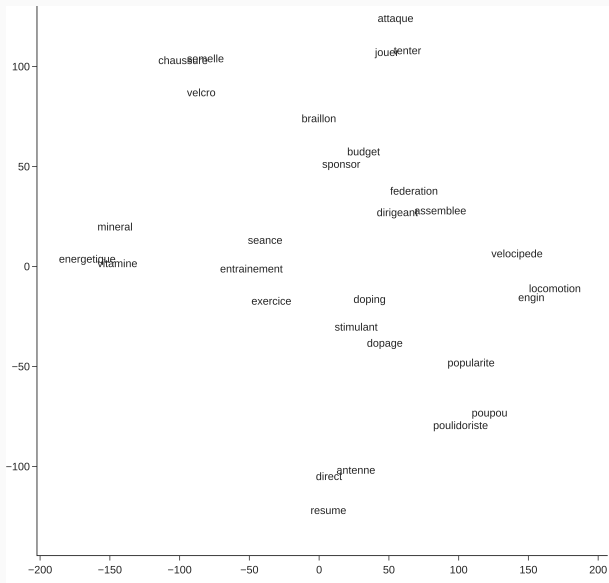


Figure – Projection (t-sne) des mots

Topic 44	locomotion	engin	velocipede	automobiliste	cyclable	circulation	automobile
Topic 89	totalisateur	journalier	affichage	chronometre	instantane	alarme	frequence
Topic 95	soufflerie	aerodynamisme	lenticulaire	aerodynamique	penetration	watt	delta
Topic 99	sucre	riz	glucide	legume	calorie	lipide	viande
Topic 5	vestige	paysage	halte	etang	siecle	detour	village
Topic 13	diametre	version	ecrou	anodiser	alliage	millimetre	acier
Topic 16	doping	dopage	stimulant	dopant	doper	repression	amphetamine
Topic 50	antenne	direct	resume	radio	filmer	emission	retransmettre
Topic 63	energetique	mineral	vitamine	glucidique	apport	boisson	nergisport
Topic 122	coffret	tome	recueil	luxueux	vpc	bibliotheque	mademoiselle

Table – Quelques topics sur les 169 trouvés

On remarque une grande variété dans les topics, même si tous ne sont pas aussi significatifs

TOPIC MODELLING : PROJECTION

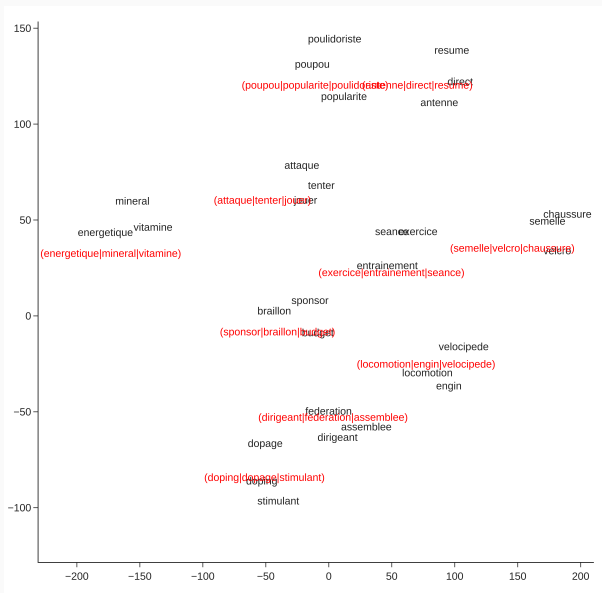


Figure – Projection (t-sne) des mots et topics

TOPIC MODELLING SUR LE SOUS CORPUS DES ÉDITORIAUX

	Topic n°1	Topic n°10	Topic n°73	Topic n°16	Topic n°0	Topic n°18	Topic n°44	Topic n°22	Topic n°115	Topic n°3
rest n°0	dégustant	peupou	descente	doping	attaque	épouse	locution	échassier	allégraph	bébé
rest n°1	lédération	populaire	ly	dépage	lettre	brûler	engin	indian	atlas	normand
rest n°2	assemblee	politiques	perle	stimulant	jeu	budget	véhicule	patenté	cas	compromett
rest n°3	politique	conscience	maître	député	patenté	retombe	autoroute	bagno	pourcentage	circumstance
rest n°4	carte	étendu	between	député	indian	invest	cybiste	ambition	massif	peupou
rest n°5	organisateur	administration	uber	repression	vainement	commanditaire	circulation	savoir	litraire	terrepreneur
rest n°6	libéral	partis	jurisou	antiracketisme	sembler	sponsoring	automobile	objectif	perle	vainement
rest n°7	instance	sincère	pellet	intéressé	litre	saire	usage	chose	indiquer	ambition
rest n°8	président	souvent	baker	sanction	démocratie	franciser	général	justement	brûler	ignorer
rest n°9	réforme	gène	marocain	si	échappée	confat	cycle	arrivé	alpes	éprouver
rest n°10	responsable	loquax	déjà	urine	terrest	modulique	véhicule	barrette	garde	pourfou
rest n°11	présidence	épouche	départ	couplet	déclencher	zapper	redorer	vainement	casu	conscience
rest n°12	déjà	déclat	franc	antidoping	déjà	démarche	danger	ingrid	graphique	confiance
rest n°13	déception	une	littor	poiff	accusation	lape	campagne	perme	tableau	égaré
rest n°14	libéral	soliste	chapit	ambassadeur	alléger	redonne	bicyclette	celer	alléger	concord
rest n°15	congrès	matérial	certain	antidoping	panache	entreprie	logica	gère	calculer	usage
rest n°16	discuss	courage	ric	médical	criser	investissement	adapte	permissible	développement	concord
rest n°17	situation	admette	poiss	médicament	équiper	communication	sième	motivation	dévoiler	docte
rest n°18	élection	jamais	curat	substance	chiquard	million	politique	cordaire	pié	stab
rest n°19	robot	condition	hété	sphérotine	gréq	employeur	instrument	quatu	parcours	soliste

Figure – Les topics les plus présents dans le corpus des éditoriaux

- On affiche les topics associés à au moins 10 documents du corpus des édits.
- On retrouve des sujets précédemment analysés tels que celui des institutions.
- De nouveaux topics intéressants émergent, comme celui sur le dopage...
- ... et également celui de l'opposition entre le vélo et la voiture.

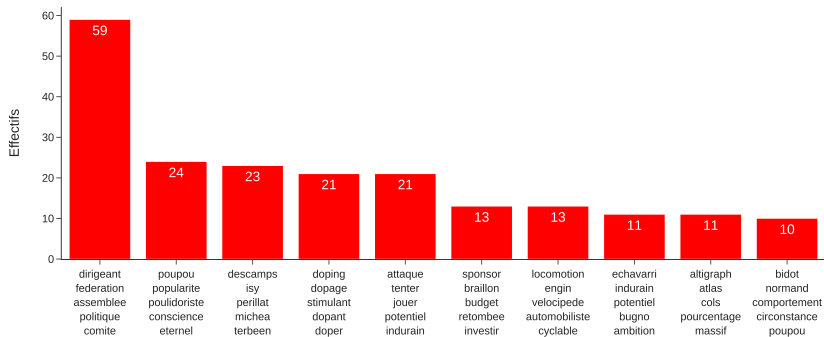


Figure – Les topics les plus fréquents dans les éditoriaux

LES FEATURES DE TOP2VEC : PROXIMITÉ DOCUMENTS-TOPICS

file_name	edito	text	topic	proximity	file_name	edito	text	topic	proximity
21	mic_1985_14_4270_2580.fr	ou	sincèrement éto delecter maurice vidal champion...	19 0.272711	247	mic_1985_11_n389_Edito.txt	ouï	sincèrement cote printemps maurice vidal trouv...	44 0.275033
76	mic_1977_14_4270_2580.fr	ou	sincèrement éto delecter maurice vidal champion...	19 0.290284					
86	mic_1982_12_4094_2580.fr	ou	maurice vidal tout armée brèche delecter suabon...	19 0.348981	271	mic_1981_04_298_Edito.txt	ouï	cyclisme fête football soldisant réalisme pren...	44 0.278682
100	mic_1977_14_4144_2580.fr	ou	affaire plongeur hère tout delecter maurice vidal...	19 0.397093	12	mic_1966_02_n068_Edito.txt	ouï	dégel février ouvrir saison cyclisme roulier s...	44 0.312844
180	mic_1985_12_4280_2580.fr	ou	sincèrement éto j'ai fait maurice vidal achulé...	19 0.397023	353	mic_1991_03_441_Edito.txt	ouï	sincèrement peleton disparu maurice vidal alim...	44 0.314417
320	mic_1988_07_4103_2580.fr	ou	sincèrement éto obéir maurice vidal tout...	19 0.411082	1	mic_1974_06_n186_Edito.txt	ouï	clauda parmentier barre industrie cycle floris...	44 0.350030
326	mic_1987_11_4083_2580.fr	ou	maurice vidal lire ce sager éditeuse cyclisme...	19 0.426266	40	mic_1963_12_n039_Edito.txt	ouï	saison roue somme vieillir petit reine année f...	44 0.367103
234	mic_1998_10_4327_2580.fr	ou	sincèrement ordinaire super maurice vidal magas...	19 0.439332	24	mic_1974-12_1975-01_n195_Edito.txt	ouï	editorial champs élysées conclusion année cycl...	44 0.411604
30	mic_1986_08_4072_2580.fr	ou	signal landi mal courreur totalement anonyme pro...	19 0.426893	57	mic_1966_03_n069_Edito.txt	ouï	drôle machine lecteur fidele écrire récemment ...	44 0.453665
136	mic_1980_08_n184_2580.fr	ou	chance passe maurice vidal non caché delecter...	19 0.433063	167	mic_1965_05_n058_Edito.txt	ouï	maurice vidal velo per coureur recemment trouv...	44 0.500991
239	mic_1977_07_08_4237_2580.fr	ou	tout probleme nio mirade norbert héri sham...	19 0.458780	196	mic_1981_10_308_Edito.txt	ouï	peur trousse français aimer bicyclette autant ...	44 0.553654
87	mic_1980_08_n115_2580.fr	ou	affaire maecte tout trace file aller édu...	19 0.402094	47	mic_1972_12_n165_Edito.txt	ouï	maurice vidal vieux dame vent cyclisme connaît...	44 0.601602
281	mic_1988_08_4140_2580.fr	ou	événement peupier ignorerne maurice vidal édu...	19 0.502151	127	mic_1973_11_n179_Edito.txt	ouï	maurice vidal cycliste dinanche pouvoir contes...	44 0.636048
112	mic_1978_12_4261_2580.fr	ou	affo pipas borned thémone borned hémone...	19 0.522354	183	mic_1974_05_n185_Edito.txt	ouï	maurice vidal million indésirable ancien lecte...	44 0.682381
218	mic_1987_08_4080_2580.fr	ou	delecter rumeurs tout étoil départ hère cartonné...	19 0.511083					
186	mic_1971_08-08_n140_2580.fr	ou	mélus incoherence france belgique sélection c...	19 0.523485					
39	mic_1976_12_4225_2580.fr	ou	clauda parmentier détonateur affaire dard par...	19 0.525547					
94	mic_1968_14_n078_2580.fr	ou	doping tout trace vitale sport spectacle ou...	19 0.588539					
260	mic_1973_11_n183_2580.fr	ou	maurice vidal chaise sonnées eddy marché dopa...	19 0.588257					
236	mic_1987_08_4094_2580.fr	ou	delecter sager jacques anquet commander ats...	19 0.608509					
23	mic_1980_07_4119_2580.fr	ou	parir longévité lation cyclisme équipé expé...	19 0.622469					

(a) Topic 16

(b) Topic 44

Figure – Les documents les plus similaires à aux topics 16 et 44

Dix millions d'indésirables

Les plus anciens lecteurs peuvent attester que le renouveau de succès de la bicyclette n'a pas surpris ce journal. Même au temps où il était de bon ton et semblait évident de proclamer qu'il s'agissait d'un engin démodé, nous n'avons cessé de répéter que la perfection mécanique ne se démode pas. Une preuve parmi d'autres ? C'est en 1966, au printemps que « Miroir du Cyclisme » intitulait un numéro « Vive le vélo ». Et je ne résiste pas au plaisir de reproduire ici quelques extraits d'un éditorial qui a plus de huit ans d'âge :

«...aujourd'hui les symptômes venus du monde entier, et très perceptibles en France, affirment que la merveilleuse petite bicyclette, cet engin soi-disant démodé mais indispensable, retrouve une grande faveur. N'est-ce pas ce que nous annoncions depuis longtemps...»

«...la bicyclette a résisté à tous les moyens mécaniques de locomotion, parce qu'elle est elle-même une grande invention, la plus simple et la plus parfaite de toutes. Elle reste le meilleur dispensateur de joies naturelles, et un outil remarquable au service de la santé.»

«Cela, cher lecteur, ne peut cependant nous faire oublier ceux dont les exploits ont porté loin et haut la renommée de leur sport, ceux qui montrant chaque dimanche quelques choses prodigieuses on peut faire avec un vélo. Tout cela ne se dit pas, mais s'ajoute...»

Cela n'est pas rappelé seulement pour raviver la mémoire de quelques écologistes de plus fraîche date, mais parce que le renouveau est venu avec une force que n'osaient imaginer les plus optimistes. Il est vrai que dans le même temps l'environnement des hommes s'assombrissait dans

des proportions que même les plus pessimistes ne pouvaient craindre.

Les jeunes revendiquent la bicyclette comme un instrument de plus grande liberté et de fête sans contrainte, les moins jeunes comme outil de détente, le meilleur allié d'une santé de toute part atteinte. Il y a en France un parc de 10 millions de vélos. Ça fait beaucoup de monde, non ? Et ça devrait faire un fameux groupe de pression, comme on dit aujourd'hui.

Eh bien, les cyclistes sont l'objet du plus grand mépris de la part des Pouvoirs Publics. Il est vrai qu'ils ne paient pas de vignette-impôt. Il est vrai qu'ils ne contribuent pas à enrichir les riches compagnies pétrolières, ni les sociétés chargées de la construction des autoroutes (et de percevoir les péages).

Mais dix millions de citoyens ! n'ont-ils pas des droits comme les autres ? Le droit de se promener dans quelques parcs ou quelques forêts sans être houlcoulés par les motos en excès de vitesse ? Le droit de circuler dans les rues des villes ou sur les routes dans quelques (modestes) couleurs qui leur seraient réservées. L'article que nous avons cité ci-dessus prouve que la vague croissante des cyclistes était facile à prévoir pour des gens qui prétendaient planifier toutes les activités du pays. Dans de nombreux pays, et même aux Etats-Unis, où l'automobile est reine, il existe des pistes cyclables et des circuits touristiques aménagés. En France, dix millions de cyclistes sont considérés comme des indésirables et circulent au péril de leur vie. Il est temps qu'ils se fassent entendre. Le « Miroir du cyclisme », journal de tous les cyclistes, les y aidera.

APRES Jacques Anquetil condamné arbitrairement, c'est maintenant au tour de Désiré Letort d'être fusillé pour l'exemple et déchu de son titre de champion de France. Ni l'un ni l'autre n'ont pu présenter leur défense. Pas plus d'ailleurs que Charly Grosskost, disqualifié au championnat du monde et qui nie formellement s'être dopé.

Jacques Anquetil a « dévoilé » des vérités premières dans un journal à sensation ; Désiré Letort a reconnu prendre de l'actiphos-amphétaminé sur les conseils de son médecin ; quant à Charly Grosskost il ingurgitait depuis quinze jours du « Pressyl » sur les conseils, lui aussi, de son médecin, pour combattre une baisse de tension, un médicament que l'on donne aux « nourrissons »...

Mais la « guerre au doping est déclarée » et la Fédération a décrété la loi martiale. Elle qui connaît ces pratiques depuis plus de cinquante ans, se net soudain à frapper aveuglément et sans aucun discernement. Elle se refait une bonne conscience sur le dos des coureurs qui n'osent plus prendre un cachet d'aspirine mais qui en reviennent aux vieilles médecines qui, elles, ne laissent pas de traces.

(a) Éditorial n°185

(b) Éditorial n°91

Figure – Deux éditoriaux en lien avec les topics 16 et 44

Top2Vec nous a permis de mettre en avant de nouvelles thématiques et des nouvelles façon de les visualiser

- Mettre en lien les topics avec les années des édito
- Analyse de tous le corpus avec le word embedding
- Top2Vec sur un plus petit corpus (suppression des publicités...)

NER : UN TOUT PREMIER RÉSULTAT

Qu'avons-nous appris ?

- Nous avons rencontré des difficultés d'analyse quantitatives liées à la taille du corpus.
- Certaines thématiques ont émergé grâce à l'analyse factorielle et la modélisation de sujets.
- ...

Pour la suite :

- Utiliser des modèles de langage de type "transformers" pour analyser notre corpus grâce à l'apprentissage par transfert.
- Améliorer la découpe du corpus.
- Sélectionner le corpus en excluant ou en isolant les publicités...
- Approfondir l'analyse des entités nommées.
- Et surtout, effectuer des analyses en fonction de vos retours et centres d'intérêt.

Merci !