Exemples d'usages de l'analyse textuelle en sociologie

Le journal d'un sans domicile

- Marpsat, Maryse. 2010. "Écrire la rue : de la survie physique à la résistance au stigmate.:Une analyse textuelle et thématique du journal d'Albert Vanderburg, sans domicile et auteur de blog." Sociologie 1(1):95–120. doi: 10.3917/socio.001.0095.
 - « La méthode Alceste est due au chercheur Max Reinert. Dans la continuité de l'analyse des données développée par Benzécri, elle cherche à rendre compte de l'organisation interne d'un discours : « Le locuteur au cours de son énonciation investit des mondes propres successifs et ces lieux, en imposant leurs objets, imposent du même coup leur type de vocabulaire. En conséquence, l'étude statistique de la distribution de ce vocabulaire devrait pouvoir permettre de retrouver la trace des "environnements mentaux" que le locuteur a successivement investis, trace perceptible sous forme de "mondes lexicaux" » (Rouré & Reinert, 1993, p. 73). »



Les discours de la BCE : garantir la stabilité

- Lebaron, Frédéric. 2015. "La BCE : flexible dans la stabilité ?" *Économie et institutions* (22). doi: <u>10.4000/ei.986</u>.
 - « Cet article analyse la relative stabilité discursive et cognitive de l'institution « banque centrale » dans le contexte de crise au sein de la zone euro, c'est-à-dire surtout depuis 2010. »
 - « Le discours n'est pas conçu ici en tant qu'entité artificiellement abstraite du reste du monde social, en particulier des propriétés sociales des locuteurs, mais en tant que processus de production de pratiques langagières toujours susceptibles d'une double analyse simultanée, linguistique et sociologique ; la pratique langagière est, dans cette optique, une composante fondamentale de la production et la reproduction de l'ordre social, et en l'occurrence de l'ordre économique et politico-institutionnel ; les discours participent à la production des politiques publiques, ne serait-ce par exemple qu'à travers les mises en récit effectuées par les acteurs de ces politiques »
 - « L'intérêt de cette démarche [de l'Analyse Géométrique des Données] est de faire apparaître les grandes différences existant au sein du corpus : les premiers axes « traduisent » les plus fortes liaisons entre mots et textes (donc aussi, auteurs). Cela permet de dégager des pôles lexicaux au sein du corpus, ou encore, si l'on préfère des « surutilisations » relatives. »





Les bobos et les geeks dans la presse

- Roquebert, Corentin. 2018. "Classer Des Styles de Vie. Proximité Lexicale et Distance Sociale Dans Le Champ Journalistique." *Politiques de Communication* 10(1):55–93.
 - Quel est l'intérêt de s'intéresser aux « pratiques journalistiques » pour étudier la circulation des catégories de représentation du monde social ?
 - Quels « pôles » du champ journalistique sont discutés ?
 - Comment ces pôles sont-ils liés à l'usage des différentes catégories ?
 - Lecture jusque p. 70





Tutoriel : construire et analyser un corpus de presse

En résumé

- Identification d'une thématique d'intérêt à analyser dans la presse (nationale ou européenne) ou twitter ou médias...
 - Réflexion par rapport au sujet de recherche du mémoire
- Recherche des articles correspondant dans la base Europresse
 - Présentation des fonctionnalités du moteur de recherche
 - Analyse des résultats sur la plateforme
 - Premiers indicateurs statistiques pertinents pour l'analyse
 - Téléchargement et transformation du corpus en base de données
 - Présentation de l'outil EuroDecodeur
 - Prolongement de la description statistique du corpus
 - Présentation de l'outil Mendak et premières manipulations
- Nettoyage des données textuelles
 - Présentation des grandes possibilités de « nettoyage » du texte pour l'analyse statistique
- Analyse du corpus
 - Analyses textuelles simples
 - Nuages de mots, typicité du vocabulaire, contextes d'utilisation des mots
 - Analyses textuelles multivariées
 - Cooccurrence, structuration du corpus et classification thématique du corpus



• <u>https://www.bib.uvsq.fr/europresse</u>

 Note : « 28 connexions simultanées sont autorisées pour toute la communauté universitaire. Merci de vous déconnecter dès que vous n'utilisez plus la ressource »



Passer en « version classique » (avancée)



La page de recherche

- Possibilité de rechercher des mots-clefs :
 - Dans tout le texte, dans le titre
 - Plusieurs mots-clefs associés ou des alternatives
 - De chercher des auteurs précis
 - De définir le champ de la recherche : presse nationale, presse spécialisée, internationale, la langue, le pays...
 - La période de recherche (attention, par défaut les 30 derniers jours)

			Bonjour Bienvenue VERSAILLES SAINT QUENTIN	
	RECHERCHER	DOSSIERS	PUBLICATIONS PDF	
echerche simple Recher	che avancee F	lecherche expre	s Recherche de biographies	
lots clés dans tout le texte				
wokisme				
) et 🖲 ou 🔾 sans [woke			dans tou	le texte 🗸 🗸
🔾 et 🖲 ou 🔾 sans 🗰 wokiste			dans tour	le texte 🗸
🔾 et 🔾 ou 🖲 sans cuisine			dans tout	le texte 🗸
Aiouter un critère Supprimer le der	nier critère)
Domaine de recherche			Dériodo	
France - Presse nationale	~		Fende	
			Dans toutes les archives 🗸	
Sources Sujets	Secteurs d'activités		○ et	
1 Nom			Sources	
Provenance			Économie ou <u>Généraliste</u> et <u>Nationale</u> ou <u>Internationale</u> et <u>Français</u> et <u>Jou</u> r	<u>naux et Quotidien et F</u>
			Sujets	
🗄 Langue			Tous les sujets	
			Secteurs d'activités	
🗄 Couverture géographique				

Les sources disponibles

- Liste exhaustive des sources disponibles : <u>https://nouveau.europresse.com/webpages/sources/sourcesearch.aspx</u>
- Présentation rapide : <u>https://www.europresse.com/seminaires-et-ateliers/</u>
- Ce qui nous intéresse surtout (et qu'on va analyser) :
 - La **presse :** articles de titres de presse
 - Les télévisions et radios : contenus de chaines (présentation, pas la transcription)
 - Les réseaux sociaux, dont des articles de blog et X/twitter MAIS attention, il ne s'agit pas d'un « moissonnage » exhaustif, et les frontières du corpus ne sont pas vraiment transparentes
 - Les études et rapports & répertoires et références : des documents « d'organismes publics et d'instances gouvernementales ainsi que d'entreprises privées »

Tableau de bord de la recherche

 \equiv





- Première analyse et raffinement des résultats
- Cliquer sur les différents éléments pour affiner au besoin les résultats

Raffiner sa recherche

 Si recherche sur une période d'un an au maximum, possibilité d'utiliser les outils de raffinement de la recherche à gauche

Affinez votre recherche



Type de média 🗸 Type de média - détail 🗸 Domaine de la source V Nom de la source V Périodicité **v** Langue de la source 🗸 Continent de la source v Pays de la source 🗸 Couverture géographique v Tonalité du document 🗸 Sujets liés au document v Secteurs d'activités du document **v** Auteur 🗸 Longueur V

Télécharger le corpus défini pour l'analyser

- On veut télécharger le maximum de textes...
 - Mais Europresse n'autorise que 1000 documents au maximum
- Pour être exhaustif, essayer de limiter sa recherche à 1000 résultats...
 - Ou alors prendre les plus récents
 - Par défaut ce sont les plus »pertinents » qui sont présentés en premier : mais comment fonctionne l'algorithme ?
- Ensuite scroller l'onglet de présentation des documents
 - Être patient !
- Cocher l'ensemble des documents
 - Avec la case tout en haut
- Enregistrer la sélection
 - Dans un fichier, choisir le format HTML
 - Patienter pendant la préparation du fichier et sauvegarder sur votre ordinateur



Date du document - Du plus récent 🕚

✓ 2024-04-25 • 743 mots

Lina Sankari - ... Parmi les invités qui se succéderont à la tribune se trouvent Geert Wilders, sorti victorieux des législatives aux Pays-Bas, mais qui a toutefois échoué à former un gouvernement, le député ...

🖬 Le Figaro

À New York, la fièvre pro-palestinienne envahit le campus de Columbia

2024-04-25 · 1430 mots PDF

Jaulmes, Adrien - ... Après Harvard, Penn et Stanford voici quelques mois, Columbia est depuis plus d'une semaine l'épicentre de ce mouvement de protestations, qui condamne pêle-mêle la politique américaine de soutien ...

Aussi paru dans 🗙

🖬 Le Figaro

Repenser de A à Z l'instruction civique !

✓ 2024-04-25 · 746 mots PDF

Ferry, Luc - ... Depuis, j'y reviens, la situation n'a fait que se dégrader, de sorte que nos cours d'instruction civique sont totalement dépassés, obsolètes. Pour l'essentiel, ils se divisent en ...

Aussi paru dans 🗙

Fin des résultats

Quel intérêt au format HTML ? (plutôt que PDF)

Le Monde

Le Monde Idées, samedi 26 avril 2025 1100 mots, p. 28

Aussi paru dans 25 avril 2025 - Le Monde (site web)

Le « wokisme » est un épouvantail pour empêcher le débat sur les discriminations

L'autrice lasbelle Karsimon, le politise Alain Policar et le philosophe Jean-Yies Pranchère répondent à un collectif d'universitaires qui estimait, dans « Le Monde » le 10 avril, que leur critique de l'« obscurantisme wole » était mathonnétement réduite à un discours réactionnair leahaite Karsimon Alain Policar Lean-Yies Pranchère

Les autous du livre collectif *Race à l'obscurantisme* wolei(à paraitre aux Presses universitaires de France, le 30 avril) se mobilisent dels aprient : dans Le Monde du 10 avril, ils et elles se plaignent de voir leur critique confondue avec le poufinisme et/ou le trumpisme et affirment filerement défendre la raison, rechercher la vérité et prôver la liberté intellecuelle. Qui pourait bien s'opposer à ces nobles idéaux – que nous partageons pleinement ? Mais le maiheur est qu'on ne défend pas la liberté intellecuelle en brandissant des las immes comme la visionisme, terme controis qui ne set qu'à disquifier sance examen rationne la lutice contre le raistime et le mascultamisme.

Entendons-nous bien : rul ne nie lexistence de pathologies militantes el sectaires, refusant l'éthique de la discussion et ostracisant des interlocuteurs. Mais celles-ci, qui doivent être condamnées, sont observables dans tout le spectre politique, et non seulement chez celles et ceux qui sont impatients d'en finir avec les discriminations. Car tel est bien le noyau définitionnel de « wole » : l'hostilité aux discriminations.

Fantasme d'un ennemi absolu

Il amine que les moynes proposés pour les combatres solient absurdes : il est tout à fait inacceptable de mettre en cuase les libertés d'expression, de création ou de programmation. Mais ces dérives n'ergagent pas plus le métus des discriminations que le libéralisme n'a été engagé par le soutien à Mussolini d'un économiste libéral comme Ludwig von Mises, ou que la Teneur de 1733 n'a discrédité à jamais le rationalisme des Lumières. Des contre-révolutionnaires comme Edmond Burke et Joseph de Maistre n'ont-lis pas dénoncé les Lumières, avant la lettre, comme un déconstructionnisme raida qui de lattra presessne tenroriste ?

C'est un fait que la plupart des lieux de diffusion médiatique de l'antiwokisme ont aussi été des lieux de diffusion de la compleisance à l'égard de Vladimir Poutine. Un des signataires de la tribune dans *Le Monde* a mené ses campagnes antiwoke dans des revues – ceiles de Michel Ontray et de la nouveile droite – qui promevent activement la pire propagande poutrineme. Ceux des antiwoke qui se targent à bon droit d'avoir toujours combatu le poutrisme ou le turnpisme occultent à trop bon compte o e qu'à été, de manière massive, le vértable sens politique de l'antiwokisme : la construction d'une « panique illibérale » par une pratique de l'antaigame généralisé. Or, orite pratique est un des modes de la « postvérité » : elle remplace l'étude scrupuleuse de la complexité des positions par la production finatsmatique d'un ennemi abolu qu'il fauited étadique. On evoit d'allemps es en qui qui la telbre entende par « volkisme » autre chose que ce qu'entendent Poutine et Trump.

Mépris du réel

Le mépris du réel se relête dans les affirmations gratuites et outrancières de la tribune. On y lit que la présomption d'innocence vaudrait suspicion de sexisme. Passons sur le fait qu'il est constitutif de la liberté d'expression qu'on puisse être critiqué pour ce qu'on dit ; passons aussi sur l'ignorance du sens de la notion de présomption d'innocence que laisse soupçonner ce propos.

La plainte fantasmatique ne sert ici qu'à faire comme si la vivacité des débats publics était un danger (qu'il faudrait combattre par la censure ?) et un problème plus grave que l'impunité du viol et l'existence des inégalités genrées : la persistance des écarts salariaux,

- On peut consulter le corpus en l'ouvrant dans un navigateur web
- Le HTML (HyperText Markup Language) est un langage de balisage
 - Il y a des balises spécifiques pour le titre, l'auteur, la date de publication, etc
 - On va se servir des balises pour extraire le corpus dans une base de données pour l'analyse...

69 70 71	L'autrice Isabelle Kersimon, le politiste Alain Policar et le philosophe Jean-Yves Pranchère répondent à un collectif d'universitaires qu //div>
72	en class-lem-margin-bettemNeuel-Trabelle Kersimon Alain Palicar Jaan-Vuer Pranchère/(n.
73	ctass= sim-inargin=bottoinvews >isabette Kersinon Atain Pottan Jean=ives Pranchere
74	
76	
77	
78	
79	
80	
81	
82	<section></section>
83	
84	div class="DocText clearfix">
85	
86	<pre>color class= ducluaried sign= s</pre>
87	sples auteurs du tive cottectif sem styte- font-styte- font-style font-s
80	
90	xx=Entendons=nous bien : nul ne nie l'existence de nathologies militantes et sectaires, refusant l'éthique de la discussion et ostracisant des interlocuteurs. Mais celles-ci, qui doivent é
91	
92	
93	
94	
95	
96	
97	
98	<aside></aside>
99	
100	
101	
102	all in
103	
109	Addition and a state of the structure structure structure of the structure structure structure structure of the structure structure of the structure
106	~/ VAY-
107	
108	

La base de données qu'on veut

- Une base où
 - les lignes sont les documents (ici les articles)
 - les colonnes sont les « métadonnées » (comme la date de publication, le nom de la source, l'auteur etc)

Le Monde		Journal 💧	Titre 🔶	Chapo 💧	Date 💧	Auteur	Texte	TypeSource 🌢	CodeSource 🌢
Le Monde Idées, samedi 26 avril 2025 1100 mots, p. 28									
Aussi paru dans 25 avril 2025 - Le Monde (site web)		All	All	All	All	All	All	All	All
Le « wokisme » est un épouvantail pour empêcher le depar sur les discriminations L'autrice Isabelle Kersimon, le politiste Alain Policar et le philosophe Jean-Yves Pranchère répondent à un collecu. Le formitaires qui estiman, dans « Le Monde » le 10 avril, que leur critique de l'« obscurantisme woke » était malhonnêtement réduite à un discours réactionnaire			Le ' wokisme ' est un	L'autrice Isabelle Kersimon, le		Isabelle	Les auteurs du livre		
Isabelle Kersimon Alain Policar Jean-Yves Pranchère Les auteurs du livre collectif <i>Face à l'obscurantisme woke</i> (à paraître aux Presses universitaires de France, le 30 avril) se mobilisent dès à présent : dans <i>Le Monde</i> du 10 avril, ils et elles se plaignent de voir leur critique confondue avec le poutinisme et/ou le trumpisme et affirment fièrement défendre la raison, rechercher la vérité et prôner la liberté intellectuelle. Qui pourrait bien s'opposer à ces nobles idéaux – que nous partageons pleinement ? Mais le malheur est qu'on ne défend pas la liberté intellectuelle en brandissant		Lewondo	époiventail pour empêcher le débat sur les discriminations	politiste Alain Policar eric philosophe Jean- Y.co: Prenchère réponde	avril 2025	Kersimon Alain Policar Jean-Yves Pranchère	l'obscurantisme woke(à paraître aux Presses universitaires de	news	LM
des leurres comme le « wokisme », terme confus qui ne sert qu'à disqualifier sans examen rationnel la lutte contre le racisme et le masculinisme.							La vie intellectuelle		
Entendons-nous bien : nul ne nie l'existence de pathologies militantes et sectaires, refusant l'éthique de la discussion et ostracisant des interlocuteurs. Mais celles-ci, qui doivent être condamnées, sont observables dans tout le spectre politique, et non seulement chez celles et ceux qui sont impatients d'en finir avec les discriminations. Car tel est bien le noyau définitionnel de « woke » : l'hostilité aux discriminations.	2	Le Monde	Les paradoxes de l'antiwokisme	Les paradoxes de l'antiwokisme	avril 2025	Livre Nicolas Truong	ressemble parfois au monde de George Orwell, qui écrivait dans 1984 : ' La	news	LM
Fantasme d'un ennemi absolu							gue		
Il arrive que les moyens proposés pour les combattre soient absurdes : il est tout à fait inacceptable de mettre en cause les libertés d'expression, de création ou de programmation. Mais ces dérives n'engagent pas plus le refus des discriminations que le libéralisme n'a été engagé par le soutien à Mussolini d'un économiste libéral comme Ludwig von Mises, ou que la Terreur de 1793 n'a discrédité à jamais le rationalisme des Lumières. Des contre-révolutionnaires comme Edmond Burke et Joseph de Maistre n'ont-ils pas dénoncé les Lumières, avant la lettre, comme un déconstructionnisme radical qui était par essence terroriste ?	3	Le Figaro	Le groupe affirme ' lutter contre les biais ' politiques de ses IA		avril 2025	Lentschner, Keren	C'est la dernière profession de foi de Mark Zuckerberg.	news né	LF
C'est un fait que la plupart des lieux de diffusion médiatique de l'antiwokisme ont aussi été des lieux de diffusion de la complaisance à l'égard de Vladimir Poutine. Un des signataires de la tribune dans <i>Le Monde</i> a mené ses campagnes antiwoke dans des revues – celles de Michel Onfrav et de la nouvelle droite – qui promeuvent activement la pire propagande poutinienne. Ceux des antiwoke qui							le fact-checking et		

Un outil pour le faire automatiquement

TRAITER LE FICHIER

去 TÉLÉCHARGER EXCEL

- EuroDecodeur
 - Une application disponible en ligne : <u>https://analytics.huma-</u> <u>num.fr/Mathieu.Ferry/EuroD</u> <u>ecodeur/</u>
 - <u>https://mathieufer.shinyapps</u> .io/europressedecodeur/
- Attention : une application encore en phase de test...
 - Marche correctement pour les corpus de presse, moins bien pour les autres types de corpus
 - Si bug, m'envoyer un email !

	Décodeur Europresse TRANSFORMATION À PROPO	9S
	Charger un fichier Europresse HTML	Tableau des articles extraits
D	BROWSE No file selected	
		Résumé par type de source détecté
<u>)S</u>	Enlever les doublons éventuels (articles très similaires)	
	Afficher le type de source détecté	
	Format de la date :	
	Conserver la date 🔹	

🕁 TÉLÉCHARGER CSV

Transformer le corpus en une base de données

Décodeur Europresse Transformation à propos

- 1. Charger le fichier HTML
- 2. Choisir si on enlève les doublons éventuels
 - Prend plus de temps !
 - Mais dans la presse, quand mise à jour d'un article sur le site, l'article apparait plusieurs fois
 - Aussi problème du format papier + format web...
- 3. Laisser cocher le type de source détectée (si on en a plusieurs différentes)



Tableau des articles extraits

Résumé par type de source détecté

4. Date: Choisir si on souhaite conserver la date de publication telle quelle (a priori pas pertinent), si on a un corpus annuel, garder mois jour, sur un mois, le jour...

Télécharger la base de données extraite par l'appli

• En csv et excel, on n'est jamais trop prudent

