

DHUM25A43 - 3

Investigating with AI

NLP

Feb 11th, 2025

What we saw last time

- Open source vs closed source
- API : REST, JSON
- Demo Colab

Projects

Design of Empirical Research with Data Analysis in Social Sciences

Defining the Research Problem : clear research questions and hypotheses

Selecting Data Sources : public datasets, archives, online platforms

Data Collection Techniques : find existing datasets, web scraping, APIs, online platforms

Data Selection, Cleaning and Preprocessing : query, data normalization and transformation

Exploratory Data Analysis (EDA) : Inspect data types, distributions, and summary statistics

- Histograms, box plots, and scatter plots for variable distribution

- Heatmaps and correlation matrices to detect relationships

- Clustering for grouping similar observations

- Mapping of data

Deeper Final Analysis: data selection, cleaning and refinement

Advanced Techniques :

- Natural Language Processing (NLP): Sentiment analysis, topic modeling, text classification

- Data Analysis with LLM

- Machine Learning Models

- Graph Theory: Social network analysis, community detection, mapping

Interpretation and Reporting : findings back to theoretical frameworks and research questions

- Data & method limitations, potential biases, and methodological constraints

Communicating Research Findings :

- Visuals and storytelling techniques, interactive charts, website

- Limitations, and future research directions

Today

- Classic NLP
 - some libraries: spacy, NLTK
- Similarity and Embeddings
- Demo : sentence similarity with wikipedia
- OpenAI API
-

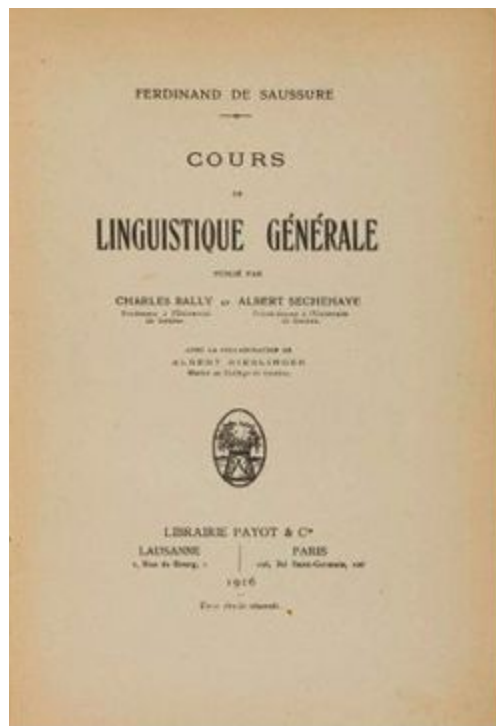
At the end of this class

You

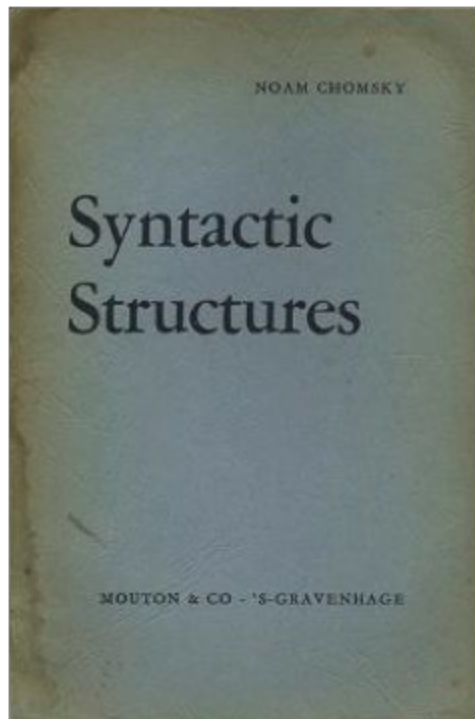
- understand what classic **NLP** is all about
- can work with embeddings
- can query OPEN AI API

Classic NLP

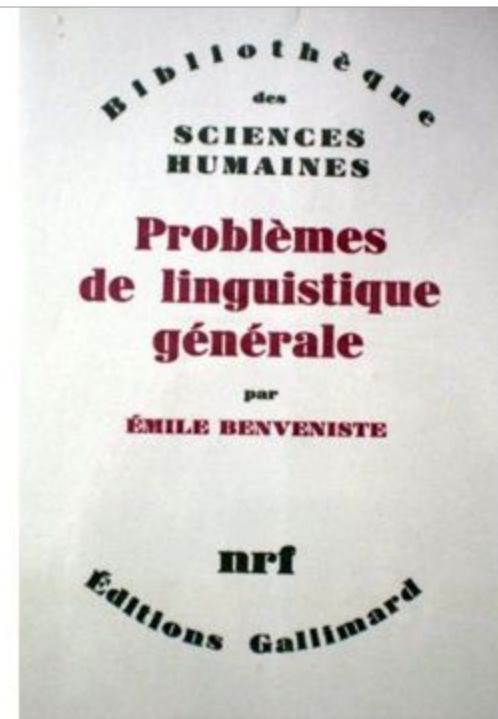
From linguistique to NLP



1916 Ferdinand de Saussure

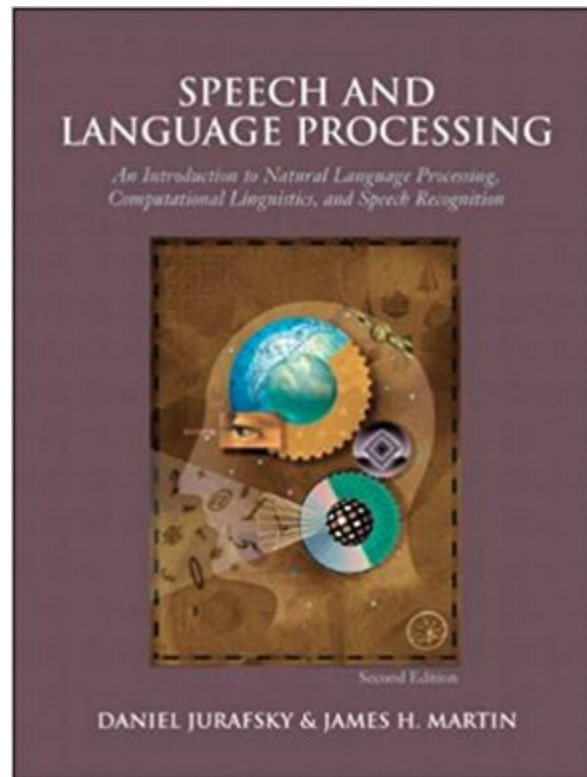


1966 Chomsky



1957 Benveniste

Book of reference



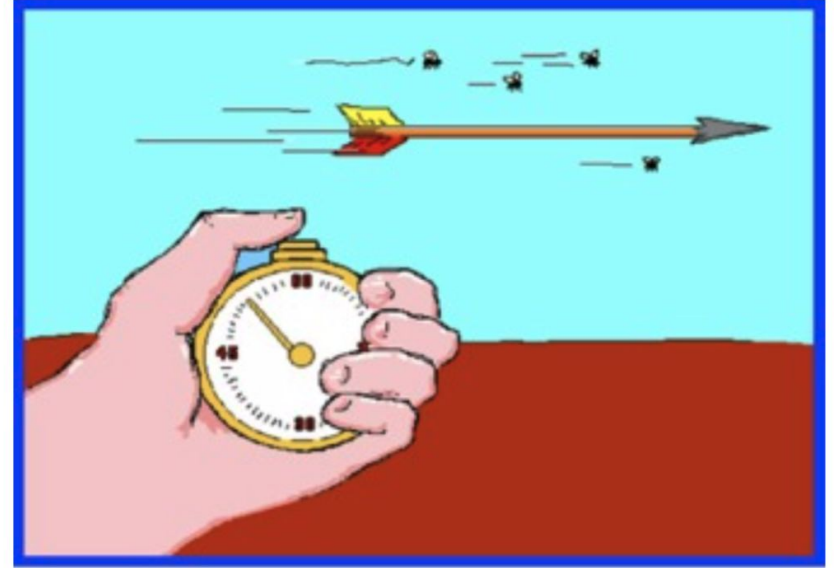
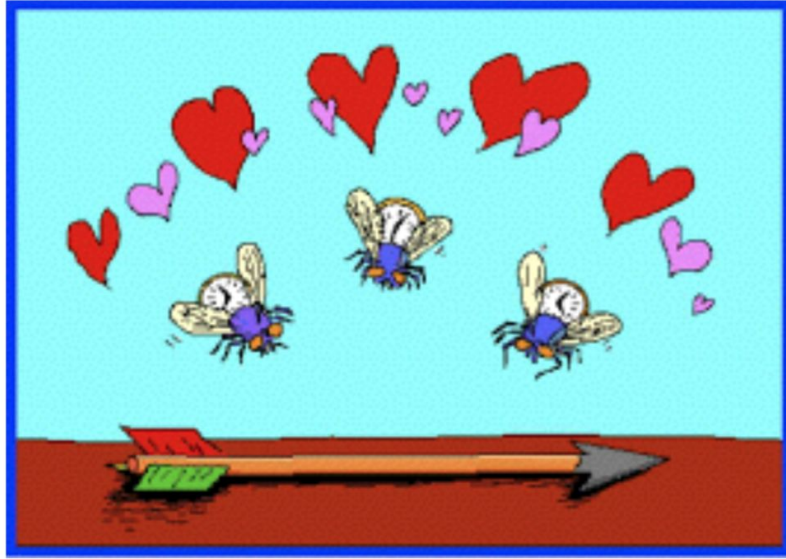
[Speech and Language Processing \(3rd ed. draft\) Dan Jurafsky and James H. Martin](#)

Ambiguities

- "Will you marry me?" is a marriage proposal.
- "Will, You, Mary, Me" is a card game proposal.
- "Will, you marry me" is a time traveller spoiling the future.
- "Will you, Mary me" is a cavewoman named Mary, trying to make Will, who has amnesia, remember who he is.

- I saw her duck
- I saw her duck. She tried to avoid a flying ball.

- Let's eat, grandpa.
- Let's eat grandpa.



Time flies like an arrow

Applications - classic NLP

- Text mining
 - NER: Named Entity Recognition: LOC, PER,
 - POS: Part of speech Tagging : nouns, adjectives, verbs, ...
- speech to text, text to speech
- Translation
- Classification: sentiment analysis, spam, ...
- Text generation, summarization, question answering
- Topic identification, topic modeling
- word sense disambiguation : bank,

Written text is extremely complex

- Non standard language
- Variable length
- Discrete: rat <> sat, while image and sound are continuous
- Wide variety of complexity across languages
 - Arabic: ...
 - German: Donaudampfschiffahrtsgesellschaftskapitän (5 “words”)
 - Chinese: 50,000 different characters (2-3k to read a newspaper)
 - Japanese: 3 writing systems
 - Thai: Ambiguous word boundaries and sentence concepts
 - Slavic: Different word forms depending on gender, case, tense
- Encoding: unicode vs Ascii
- **unstructured data**

Wayyyyyyy back in 2021! 4 years ago!



Easy

- Chunking
- Part-of-Speech Tagging
- Named Entity Recognition
- Spam Detection
- Thesaurus



Medium

- Syntactic Parsing
- Word Sense Disambiguation
- Sentiment Analysis
- Topic Modeling
- Information Retrieval



Hard

- Machine Translation
- Text Generation
- Automatic Summarization
- Question Answering
- Conversational Interfaces

Demo classic NLP

<https://cloud.google.com/natural-language>

Valery Gerasimov, chief of the army's general staff, said the Tsirkon missile covered a distance of more than 280 miles in 4.5 minutes to destroy a target in the Barents Sea off Russia's Arctic coast. He said it was fired from an Admiral Gorshkov-class frigate in the White Sea and that it marked the first time the missile had hit a target at sea.

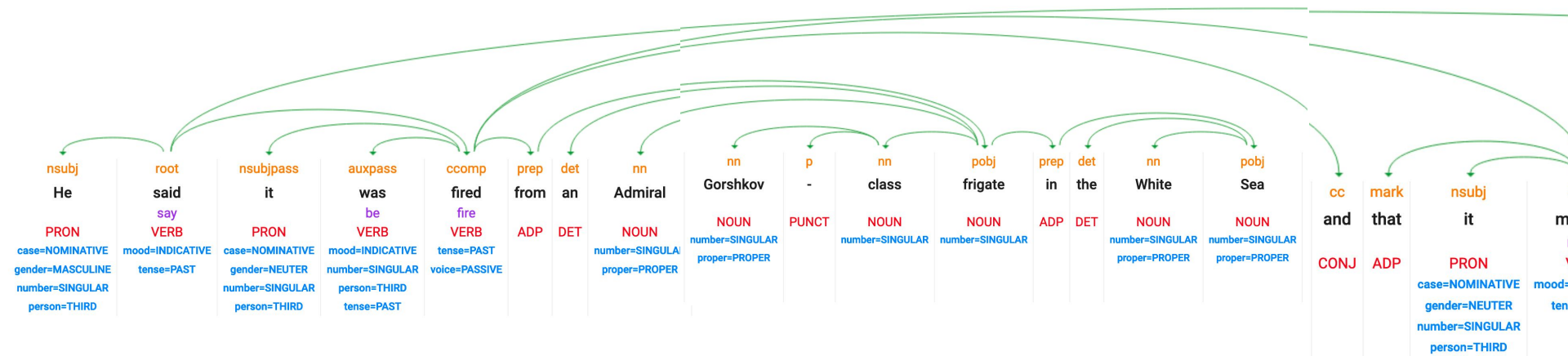
From [Russia tests hypersonic missile on Putin's birthday | World](#)

SpaceX began a new era of spaceflight: For all those searching for a new planet to call home, the year brought at least a bit of good news. SpaceX, the company founded by billionaire Elon Musk to fulfill his dream of colonizing Mars, launched NASA astronauts into orbit for the first time since the U.S. government retired the space shuttle program in 2011. SpaceX regularly transports cargo to the International Space Station, and in 2020 became the first private enterprise ever to launch astronauts there.

From [2020 Events - Pop Culture, U.S. Politics & World | HISTORY](#)

Part of Speech tagging

✓ Dependency ✓ Parse label ✓ Part of speech ✓ Lemma ✓ Morphology



Stopwords

- many words have no meaning: stopwords

```
stopwords = ['the', 'of', 'and', 'is', 'to', 'in', 'a', 'from', 'by', 'that',  
             'with', 'this', 'as', 'an', 'are', 'its', 'at', 'for', ...]
```

What is the unit ?

We can work with

- words, syllables, tokens, letters & punctuation
- **Bigrams, n-grams**: New York, cul-de-sac, pain au chocolat
- sentences, paragraphes, tweets, articles, books, comments
- Corpus: a whole set of text

And

- The vocabulary is large / infinite and fast changing
- Typos, multiple spellings

Vocabularies

- fixed and large size : balance between performance and exhaustivity / meaning
- Out of Vocabulary
 - Named entities and proper nouns
 - Domain-specific technical terms
 - Neologisms and emerging words
 - Misspellings and typos
 - Morphological variants

Tokenization: defining the unit of text

- What's the most efficient unit of text ?
- Using words is problematic:
 - very large vocabulary, multiple forms: plural, declension (home, house), conjugation, ...,etc
- Using letters is too short

Let's split words into multiple tokens: **subwords** tokenization

- unhappiness -> un + happ + iness
- running -> run + ing
- biiig -> b + iii + g

Benefit of subwords tokenization

- no words are OOV: out of vocabulary
 - gpt 3.5 did not know about COVID
- typo resilient
- smaller vocabulary size
- captures semantic and morphological meaning better

LLM pricing - based on input - output Tokens

OpenAI o1

Frontier reasoning model that supports tools, Structured Outputs, and vision | 200k context length

Price

Input:

\$15.00 / 1M tokens

Cached input:

\$7.50 / 1M tokens

Output:

\$60.00 / 1M tokens

OpenAI o3-mini

Small cost-efficient reasoning model that's optimized for coding, math, and science, and supports tools and Structured Outputs | 200k context length

Price

Input:

\$1.10 / 1M tokens

Cached input:

\$0.55 / 1M tokens

Output:

\$4.40 / 1M tokens

[Pricing | OpenAI](https://openai.com/pricing) Feb 4th 2025

Embeddings

A distance between words

How to calculate a distance between 2 texts ?

so that we can say that

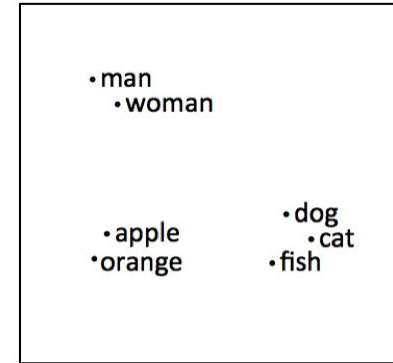
- **banana** is closer to **apple** than it is to **plane**;
- the **dog barks** is closer to the **cat meows** than it is to the **tractor sputters**, ...

before 2013: counting words and their relative frequencies
(messy, crude, called tf-idf, worked well for spam detection)

Word2Vec

2013: Thomas Mikolov @Google Efficient Estimation of Word Representations in Vector Space <https://arxiv.org/abs/1301.3781>

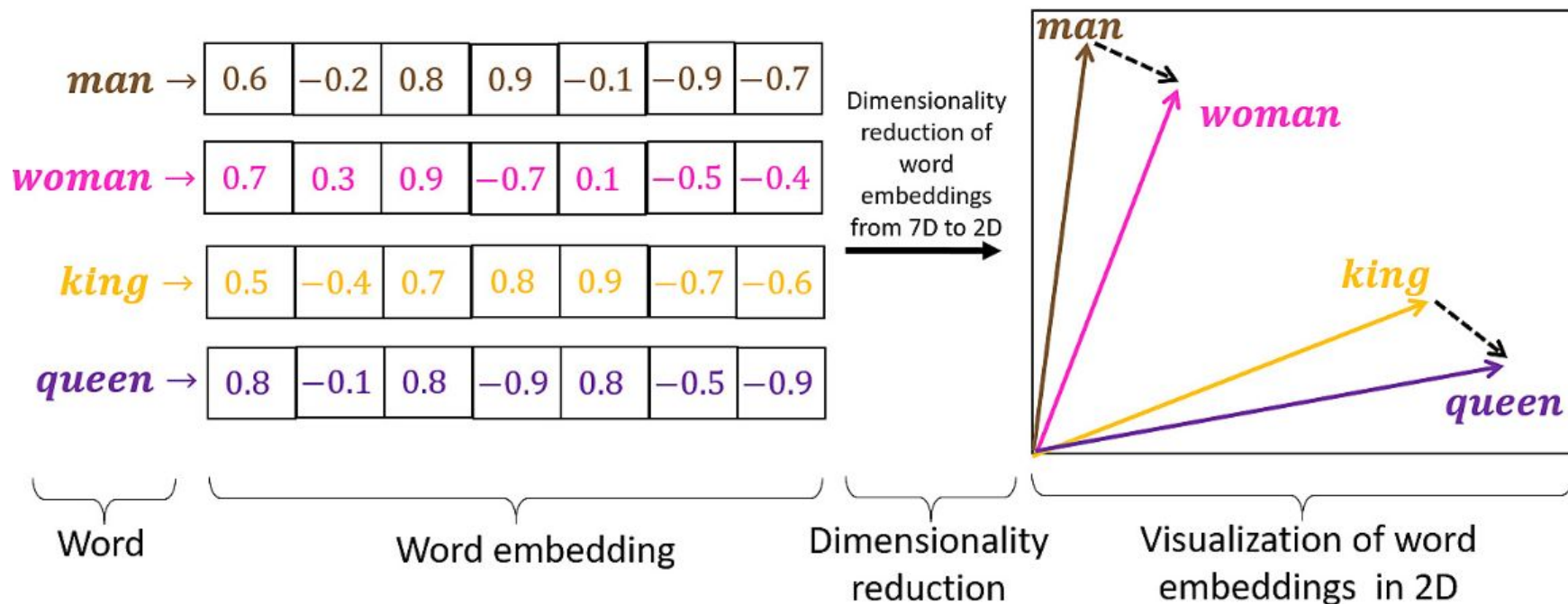
- Queen - Woman = King - Man
- France + capital = Paris



no context disambiguation: bank (river) =
bank (finance)

2013! Word2vec

One word has one vector of large dimension



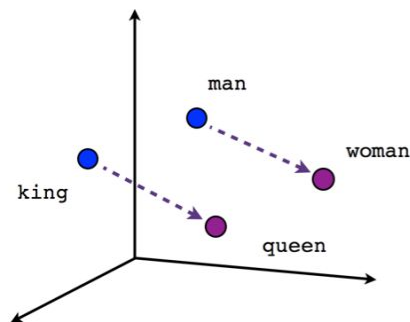
<https://airbyte.com/data-engineering-resources/sentence-word-embeddings>

distance with Cosine Similarity

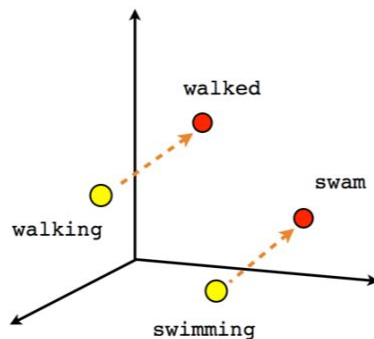
Once you have embeddings of 2 texts (sentences, words, tokens ,...)

The distance between the 2 text is given by the cosine similarity of their embeddings

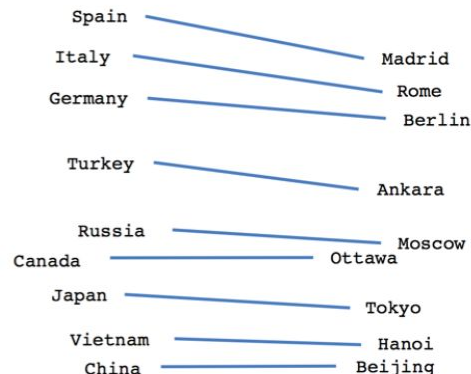
$$\text{sim}(u, v) = \frac{u^T v}{\|u\|_2 \|v\|_2}$$



Male-Female



Verb tense



Country-Capital

Current embeddings methods

Embedding using an LLM

Hugging Face Transformers: For models like BERT, RoBERTa, SBERT.

OpenAI API: For models like text-embedding-ada-002.

Example Workflow:

1. Load a pre-trained SBERT model.
2. Pass your sentences through the model.
3. Get a 768-dimensional vector (embedding) for each sentence.
4. Use these embeddings for your task.

Demo

Find immigration related text in wikipedia pages of main European capitals

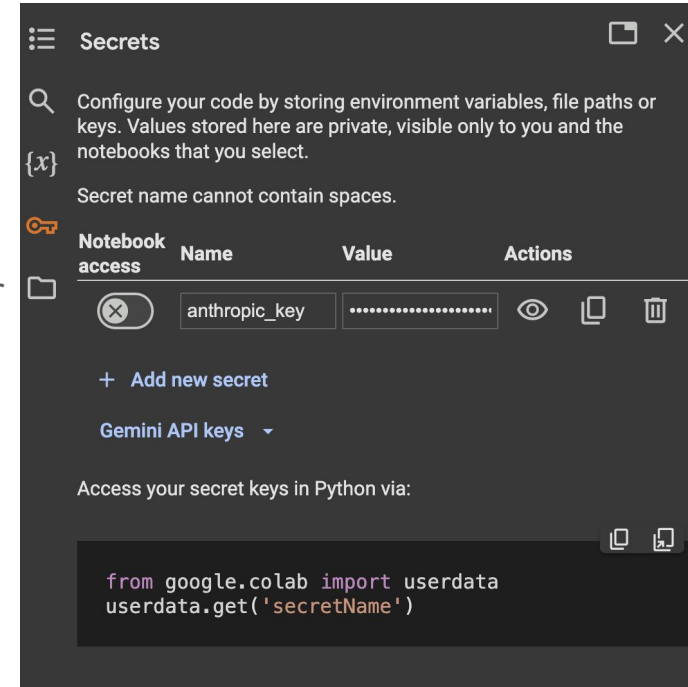
- use chatGPT project
- use wikipedia API
- get the pages for Paris, Rome, Berlin, ...
- split the text into sentences
- calculate the embeddings for each sentence
- given a query sentence and its embeddings
 - find the N sentences that are most similar to the query sentence
- save the corpus

<https://colab.research.google.com/drive/16wQ3o8XQNEhADR2v2ibwfvjaKXKvkyKW#scrollTo=cGm37GYrSNd9>

OpenAI API

The **secret** key

- create an account
- provision some money (~\$2, \$3)
- create a **secret** key
 - save that secret key in a safe place on your computer
 - never put the key in the code
 - in colab, add the key on the left side
- for OpenAI the name of the key is
 - OPENAI_API_KEY



Data Sources

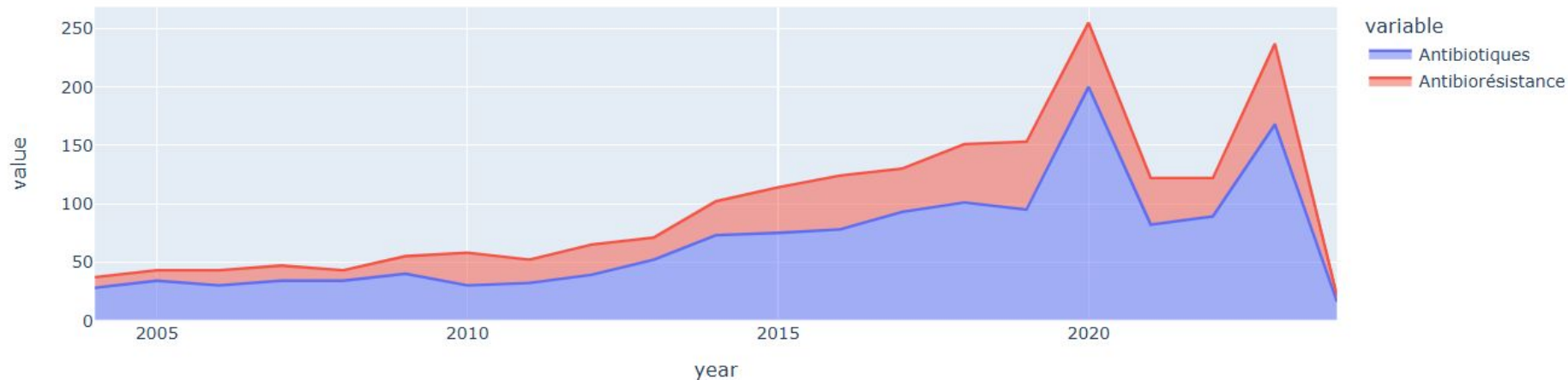
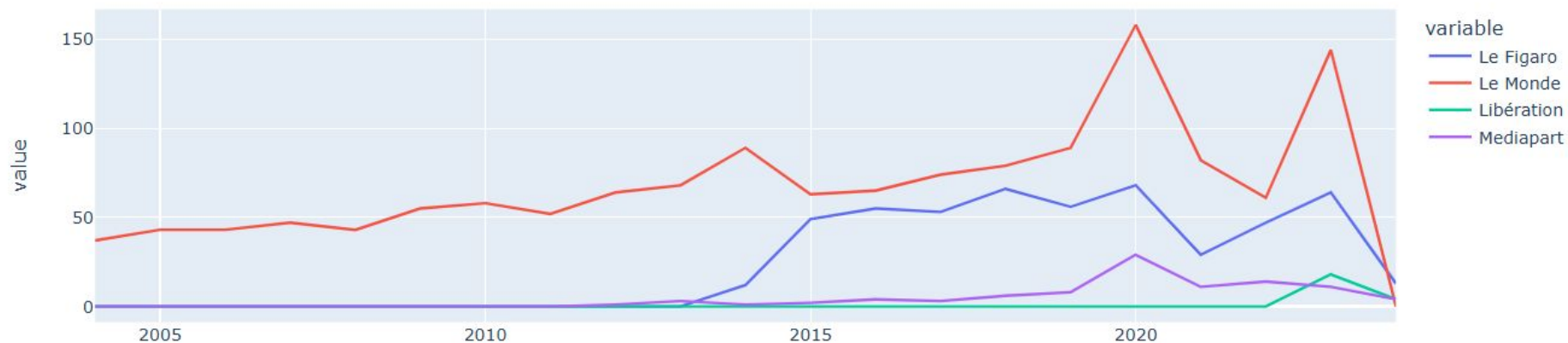
- Doctissimo (forum “antibio”)
- Twitter (-> 2023)
- Youtube
- Presse ('Mediapart', 'Le Figaro', 'Libération', 'Le Monde') (+ Factiva)
- Web

Query

((resistant OR resistente OR resistance OR résistantes OR resistance OR résistant OR résistante OR résistance OR résistantes OR résistance) AND (antibiotique OR antibiotiques OR antibio OR antibios OR antimicrobien OR antimicrobiens)) OR antibioresistance OR antibiorésistant OR antibioresistente OR antibiorésistance OR antibiorésistant OR antibiorésistante OR #antibioresistance OR #antibiorésistant OR #antibioresistente OR #antibiorésistance OR #antibiorésistant OR #antibiorésistante)

((resistant OR resistente OR resistance OR résistantes OR resistance OR résistant OR résistante OR résistance OR résistantes OR résistance) (Amphotericine OR Amikacine OR Gentamicine OR Neomycine OR Nétilmycine OR Spectinomycine OR Streptomycine OR Tobramycine OR colistine OR Polymyxine OR Daptomycine OR Vancomycine OR Teicoplanine OR Fosfomycine OR Ceftriaxone OR Céfépime OR Céfotaxime OR Cefpirome OR Ceftazidime OR Céfixime OR Cefpodoxime OR Céfidiam OR Céfamandole OR Céfoxitine OR Céfuroxime OR Céfaclor OR Céfadroxil OR Céfalexine OR Céfalotine OR Céfazoline OR Céfradine OR Aztréonam OR Ertapénem OR Imipénem OR Méropénem OR Témocilline OR Pivmécillinam OR Pipéracilline OR Tazobactam OR Cilastatine OR Ticarcilline OR (Acide clavulanique) OR Oxacilline OR Cloxacilline OR Pénicilline OR Benzathine OR benzylpénicilline OR Ampicilline OR Sulbactam OR Amoxicilline OR cefpodocile OR proxetil))

Antibiotics in 'Mediapart', 'Le Figaro', 'Libération', 'Le Monde' 2004-2024 (2045)



LLM Prompt

```
data_analysis_task = Task(
    description="""Analyze textual data if French {text} as a sociologist and provide
    1. extract a list of actors (actants) without duplicates, categories of actants with examples is listed below
    2. for each actor (actant) extract logics, worlds, arguments, related topics
    3. for each actor (actant), logics, worlds provide categorisation based on provided categories

Output the results in French.

STRICTLY Using following list of categories with lists example items:

ACTANTS:
"Agricultural actant": ["agriculteur", "agriculteurs", "agricultrice", "agricultrices", "agriculture", "agricultures", "apiculteur", "apiculteurs", "apicultrice", "a
"Antimicrobial actant": ["antimicrobien", "antimicrobiens", "microbicide", "microbiostatique", "sulfamide", "antibiotique", "antibiotiques", "antibio", "antibios", "a
"Biological actant": ["covid 19", "covid-19", "parasite", "parasites", "ravageur", "ravageurs", "virus", "bactérie", "bactéries", "bactérie résistante", "bactéries r
"Chemical actant": ["adjuvant", "chimique", "chlordécone", "CO2", "co2", "composé", "COV", "gaz à effet de serre", "ges", "glyphosate", "néonic", "néonicotinoïde", "
"Economic actant": ["acteur économique", "acteurs économiques", "Bayer", "Bouygues", "conglomérat", "entrepreneurs", "entreprise", "entreprises", "Ericsson", "Free", "
"Legal actant": ["avocat", "avocate", "avocates", "avocats", "cour de justice", "juge", "jury", "magistrat", "magistrate", "magistrates", "magistrats", "tribunal"],
"Medical actant": ["médecin", "médecins", "pharmacien"],
"Physical actant": ["électromagnétique", "électromagnétiques", "faisceaux", "ionisant", "ionisants", "onde", "ondes", "radiation", "rayon", "rayons"],
"Political actant": ["barbarapomplii", "Bruno-Lemaire", "Castaner", "Christian Jacob", "de Rugy", "Denormandie", "deRugy", "FranceInsoumise", "Hulot", "J.
"Public Health actant": ["agence", "agence sanitaire", "anses", "bfr", "circ", "echa", "efsa", "IARC", "IARCWHO", "OMS", "OMSA", "OIE", "FAO", "PNUE", "Quadrupartite
"Public authorities": ["Agri_gouv", "Agriculture Canada", "Assemblée", "Assemblée nationale", "autorité", "autorités", "Bruxelles", "commission européenne", "Député"
"Stakeholder actant": ["antivax", "association", "associations", "Bayer", "collectif", "collectif", "collectifs", "complotiste", "complotistes", "CR30", "Dow", "écol
"Technological actant": ["antennes", "capteur", "capteurs", "Internet des objets", "Internet of Things", "iot", "puce", "réseau", "satellite", "serveur", "serveurs",
"Vulnerable actant": ["amish", "citoyen", "citoyens", "électrosensible", "électrosensibles", "femmes", "hypersensible", "hypersensibles", "nourrisson", "nourrissons"
"Zoological actant": ["abeilles", "animaux", "chauve-souris", "chauves-souris", "insectes", "oiseau", "oiseaux", "pollinisateur", "pollinisateurs"],

LOGICS:
"Economic logic": ["abonnement", "achat", "argent", "cher", "commercialiser", "compétitivité", "concurrence", "concurrencer", "consommer", "coût", "croître", "dette"
"Environmental logic": ["biologique", "carbone", "dégrader", "disparition", "écologique", "émettre", "environnemental", "jardiner", "menacer", "pesticide", "polluant
"Inquiring logic": ["analyse", "analysé", "analyser", "apprendre", "avéré", "cartographie", "chargé", "classe", "collecte", "collecter", "concluaît", "conclusion", "
"Judicial logic": ["attaqué", "attaquer", "condamner", "illégal", "justice", "légal", "plainte", "plainte", "poursuit", "procès", "recours"],
"Protesting logic": ["aberrant", "aberration", "anti", "argument", "battez", "battre", "bronche", "broncher", "colère", "conneries", "contre", "convaincu", "coup", "
"Regulating logic": ["adopter", "adopter", "amendement", "actes délégués", "arrêt", "arrêté", "arrêter", "autorisation", "autoriser", "autorités", "bannir"
"Revealing logic": ["accablant", "alerte", "allégation", "bâilloné", "bizarre", "blesse", "bombe", "complot", "complotiste", "corrompre", "criminels", "cynisme", "c
"Sanitary logic": ["nocif", "nocive", "nocifs", "nocives", "sanitaire", "santé", "cancer", "maladie", "contaminer", "propager", "toxique", "toxiques", "vacci
"Surveillance logic": ["accéder", "dictature", "espionnage", "guerre froide", "sécuri", "sécuriser"],

WORLDS:
"World of administration": ["arrêté", "autorisation", "autorisations", "autorité de la concurrence", "Commission", "commission", "derogation", "dérogation", "dérogat
"World of agriculture": ["champs", "CR30", "culture", "cultures", "épandage", "exploitations", "filière", "FNSEA", "herbicide", "neonic", "néonicotinoïde", "pesticide
"World of cybersecurity": ["backdoor", "information personnelle", "informations personnelles", "intelligence économique", "logiciel", "logiciel espion", "porte dérot
"World of ecology": ["biodiversité", "changement climatique", "changements climatiques", "eaux", "écologies", "écologies", "écologique", "écologiquement", "écologie
"World of economy": ["compétitivité", "concurrences", "coût", "coûts", "économie", "marché", "marchés", "production", "productions", "productivité"],
"World of health": ["alzheimer", "bien-être", "cancer", "cancers", "chirurgical", "chirurgicale", "danger", "dangereux", "empoisonnement", "endocrinien", "LMA", "mal
"World of influence": ["conflit d'intérêt", "conflit de intérêt", "corruption", "leader", "lobbying"],
"World of justice": ["avocat", "cour de justice", "cours", "magistrat", "palais de justice", "plaint", "plainte", "plaintes", "procès", "Tribunal Monsanto", "tribuna
"World of protest": ["associatif", "contestation", "genefutures", "mobilisation", "mobilisations", "opposition", "protestation", "Tribunal Monsanto"],
"World of science": ["échantillon", "échantillons", "hypothèse", "hypothèses", "laboratoire", "laboratoires", "prélèvement", "prélèvements", "preuve", "p
"World of technology": ["automatisation", "capacité", "massif MIMO", "massive MIMO", "network slicing", "radio-identification", "réalité augmentée", "réalité virtuelle
...
expected_output="""You must output only table with 6 columns: actants, logics, worlds, arguments, related topics and categorisation. Output the results in French.
```

Annotation by LLMs - actants, arguments, topics, summary

Matthieu Amiech: «La gestion de la crise sanitaire nous entraîne vers une société-machine» Pour l'éditeur Matthieu Amiech, penseur du développement des technologies, la crise liée au Covid constitue «un effet d'aubaine pour les géants du numérique». Face à «l'informatisation de toute la vie sociale», face à la mise «à l'arrêt» ou presque de la vie démocratique, les citoyens se retrouvent aujourd'hui «sans défense morale et politique». Sa réponse: une «désobéissance concertée». Voilà un an que nous vivons sous le «régime Covid». Un nouveau régime de relations sociales fait de perte du toucher, de disparition du festif, d'assèchement culturel et de reculs démocratiques. Un régime où le numérique s'impose à tous les niveaux, sans que l'impact écologique de cette mutation ne soit jamais interrogé.

Pour Matthieu Amiech, penseur de longue date du développement des technologies, celle-ci est pourtant profonde. En 2013, déjà, il alertait, dans un ouvrage collectif signé du groupe Marcuse (Mouvement autonome de réflexion critique à l'usage des survivants de l'économie), La Liberté dans le coma – Essai sur l'identification électronique et les motifs de s'y opposer, sur les menaces que fait peser la société numérique sur l'égalité et les libertés...

domains	topics	summary
milieu & nature	transition agroécologique	Matthieu Amiech critique l'impact des technologies numériques amplifié par la crise sanitaire du Covid-19, soutenant qu'une transformation vers une "société-machine" se dessine, accompagnée d'un affaiblissement des libertés et de la démocratie.
Europe & marché	industrie pharma	Il souligne que le numérique renforce le contrôle technocratique et autoritaire, entraînant une concentration des pouvoirs dans les mains de géants technologiques au détriment des pratiques démocratiques.
lutte & Organisation mondiale de la santé	pandémie - bizarre	Amiech appelle à une désobéissance concertée pour s'opposer à cette numérisation omniprésente, et il met en exergue la nécessité d'une réflexion collective sur notre rapport à la technologie et ses conséquences sur l'écologie et la société.
histoire & guerre	histoire des épidémies	La gestion de la pandémie actuelle rappelle les précédentes crises sanitaires, notamment par son approche bureaucratique et sa tendance à favoriser les intérêts industriels au détriment des enjeux environnementaux et de santé publique.
personnel & soins	recit d'hospitalisation	Amiech évoque les conséquences sociales de l'isolement imposé par la pandémie, insistant sur la déshumanisation des interactions sociales et le besoin d'un retour à des pratiques conviviales et de solidarité.

o actant_name	o logics	o worlds	o arguments	o topics_actant	o actant
matthieu amiech	protesting logic	world of protest	critique des dérives de la numérisation et appel à une désobéissance conc...	désobéissance, r...	stakeholde
géants du numéri...	economic logic	world of econ...	la gestion technocratique de la pandémie profite aux géants du numérique...	crise sanitaire, c...	economic e
citoyens	protesting logic	world of protest	les citoyens sont désormais sans défense morale et politique face à la num...	démocratie, mobi...	vulnerable
autorités publiques	regulating logic	world of admi...	la gestion de crise centralisée entraîne des reculs démocratiques et des m...	technocratie	public auth
collectifs / mouve...	protesting logic	world of protest	appel à se rassembler et à créer des mouvements populaires pour contrer l...	mouvements soc...	stakeholde
experts	economic logic	world of econ...	la course à l'innovation technologique prend le pas sur des approches dé...	technocratie, do...	political act
milieu de la santé	sanitary logic	world of health	la focalisation sur des solutions technologiques néglige d'autres approches...	vaccination, com...	public heal
ig	surveillance l...	world of tech...	la 5g favorise une surveillance intensive et distance davantage les citoyens...	contrôle, surveill...	technologic
agriculture	environment...	world of agric...	automatisation des processus dans le cadre de l'agriculture 4.0 menace l'e...	chômage technol...	agricultural
vaccin	sanitary logic	world of health	les vaccins sont présentés comme une solution miracle, sans réfléchir aux ...	crise sanitaire, d...	medical act

Annotation by LLMs - actants, arguments, topics

	actant_name	logics	worlds	arguments	topics_actant	actants	id	idl	year
0	matthieu amiech	protesting logic	world of protest	critique des dérives de la numérisation et app...	désobéissance, résistance	stakeholder actant	0	0	2021
1	géants du numérique	economic logic	world of economy	la gestion technocratique de la pandémie profi...	crise sanitaire, contrôle	economic actant	0	0	2021
2	citoyens	protesting logic	world of protest	les citoyens sont désormais sans défense moral...	démocratie, mobilisation	vulnerable actant	0	0	2021
3	autorités publiques	regulating logic	world of administration	la gestion de crise centralisée entraîne des r...	technocratie	public authorities	0	0	2021
4	collectifs / mouvements	protesting logic	world of protest	appel à se rassembler et à créer des mouvement...	mouvements sociaux, résistance	stakeholder actant	0	0	2021
...
8421	équipes hospitalières	sanitary logic	world of health	mise en place de protocoles qui ont réduit la ...	protocole, réanimation, soins	medical actant	702	702	2023
8422	ministères de la santé	regulating logic	world of administration	critiques des mesures actuelles face à la mont...	nouvelles mesures, prévention	public authorities	702	702	2023
8423	chercheurs en santé	inquiring logic	world of science	nécessité de recherches pour mieux comprendre ...	recherche, compréhension	public health actant	702	702	2023
8424	associations de patients	protesting logic	world of protest	appel à une mobilisation collective pour la se...	mobilisation, sensibilisation	stakeholder actant	702	702	2023
8425	patients	vulnerable logic	world of health	éprouvent des difficultés face à l'ignorance d...	inconnu, mortalité	vulnerable actant	702	702	2023

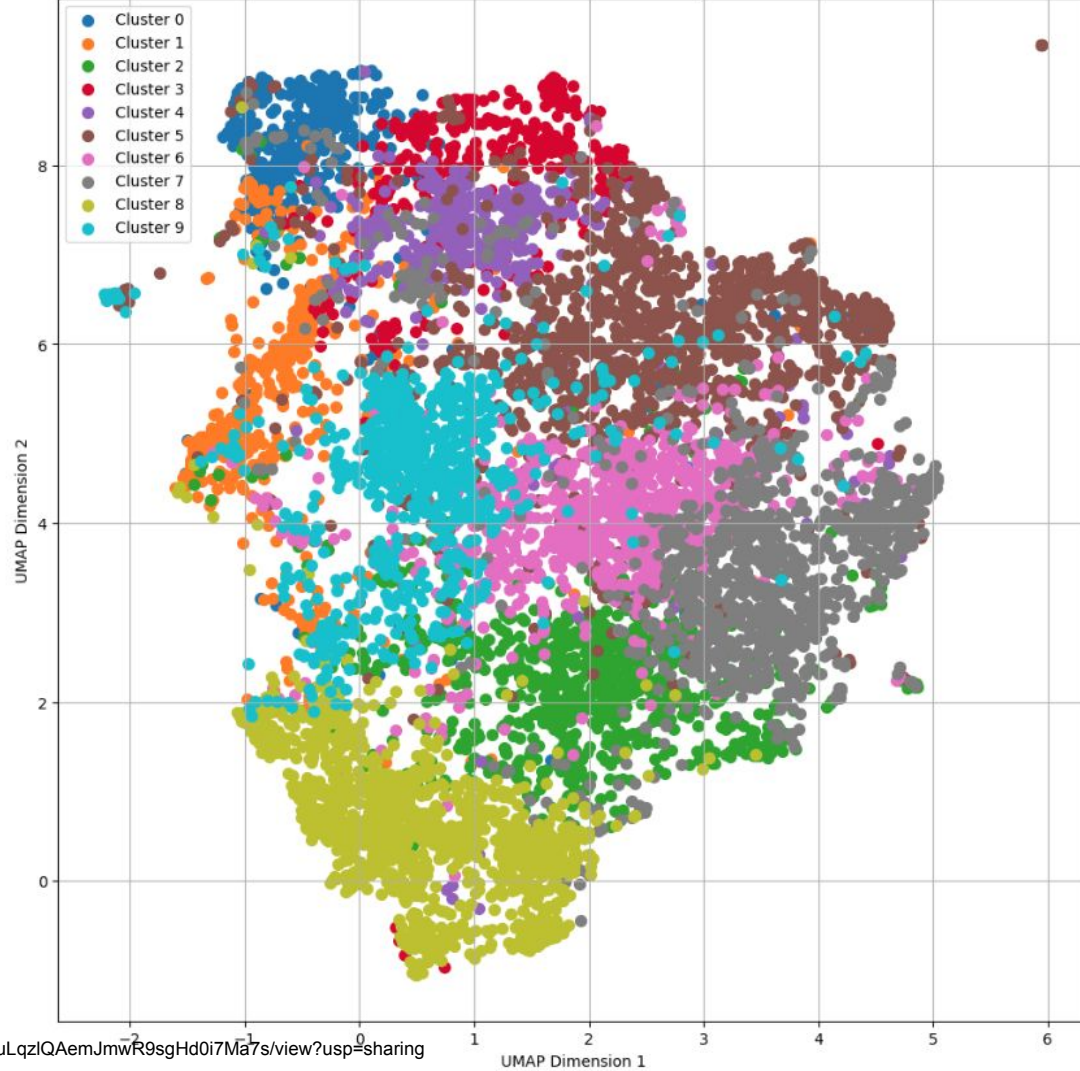
Annotation by LLMs - topics, summary

0	1	2
domains	topics	summary
milieu & nature Europe & marché lutte & Organisation mondiale de la histoire & guerre personnel & soins domains	transition agroécologique industrie pharma pandémie - bizarre histoire des épidémies récit d'hospitalisation topics	Matthieu Amiech critique l'impact des technologies numériques amplifié par la crise sanitaire du Covid-19, soutenant qu'une transformation vers une "société-machine" se dé... Il souligne que le numérique renforce le contrôle technocratique et autoritaire, entraînant une concentration des pouvoirs dans les mains de géants technologiques au détrime... Amiech appelle à une désobéissance concertée pour s'opposer à cette numérisation omniprésente, et il met en exergue la nécessité d'une réflexion collective sur notre rapp... La gestion de la pandémie actuelle rappelle les précédentes crises sanitaires, notamment par son approche bureaucratique et sa tendance à favoriser les intérêts industriels... Amiech évoque les conséquences sociétales de l'isolement imposé par la pandémie, insistant sur la déshumanisation des interactions sociales et le besoin d'un retour à des... summary
histoire & guerre milieu & nature lutte & Organisation mondiale de la cause & mesures domains	histoire des épidémies pandémie - bizarre prévention & maladies infectieuses médicaments chers - budget médicaments topics	Le texte explore l'impact historique des pandémies, en mettant l'accent sur la grippe espagnole de 1918-1919 comme parallèles de l'actuelle crise du coronavirus. Il souligne... Cet article traite des implications écologiques et sociétales des pandémies, notamment comment elles révèlent l'interconnexion entre les espèces humaines et animales. Les... L'analyse met en évidence la nécessité d'une approche organisée et globale face aux pandémies. Les experts insistent sur l'importance d'unir les efforts des médecins et vé... Le texte souligne l'impact économique des pandémies sur les systèmes de santé et les ressources nécessaires pour lutter contre les épidémies. Les chercheurs mentionner... summary
cadre & autorisation cause & mesures prévention & maladies infectieuses domains	couverture vaccinale couverture vaccinale couverture vaccinale topics	Le gouvernement français propose d'étendre l'obligation vaccinale à 11 vaccins, malgré une opposition forte et des doutes parmi les professionnels de santé quant à l'efficac... L'article aborde la question des mesures à prendre pour améliorer la couverture vaccinale en France. Le rapport de la concertation citoyenne remet en question l'obligation v... La situation actuelle de la vaccination en France est critique, avec des taux de couverture insuffisants pour certaines maladies comme la rougeole. L'article souligne que la c... summary
industrie pharma Europe & marché domains	industrie pharma médicaments chers - budget médicaments topics	Quentin Ravelli analyse la résilience de l'industrie pharmaceutique face aux scandales répétés, comme Mediator et Vioxx. À travers une enquête approfondie, il révèle comm... Le texte discute des profits continus de l'industrie pharmaceutique et de son influence politique, malgré les critiques et les nettements défallants dans le système médicame... summary
Europe & marché cadre & autorisation lutte & Organisation mondiale de la cause & mesures - domains	- - - - - topics	L'article aborde le concept de « lanceur d'alerte » développé par Francis Chateauraynaud, qui distingue cette notion des termes anglo-saxons comme « whistleblower ». Il y... Chateauraynaud explique que la notion de lanceur d'alerte concerne aussi bien des citoyens ordinaires que des professionnels en rupture de ban. Le texte souligne l'importa... Le texte examine les différents processus qui mènent à la comparaison entre alertes et mobilisations, en soulignant qu'une alerte efficace nécessite des actants qui peuvent... L'auteur aborde la complexité des systèmes de prise en compte des alertes, qui dépend de l'existence de précédents, de la nature des risques et de la compréhension socié... En somme, Chateauraynaud qualifie le lanceur d'alerte de figure centrale en démocratie, signalant que les vrais problèmes émergent souvent par des moyens inhabituellem... summary
usage & antibiorésistance cause & mesures lutte & Organisation mondiale de la domains	microbiologie industrie pharma lutte topics	Le texte met en lumière la menace croissante posée par les bactéries multirésistantes aux antibiotiques, signalant une augmentation alarmante des cas d'infections résistan... Le document souligne que malgré une baisse de la consommation d'antibiotiques en France, celle-ci est repartie à la hausse ces dernières années en raison de prescriptio... Le texte aborde l'alerte lancée par l'OMS concernant l'antibiorésistance, qui entraîne des décès évitables, soulignant l'urgence de mobiliser les gouvernements et les systèm... summary

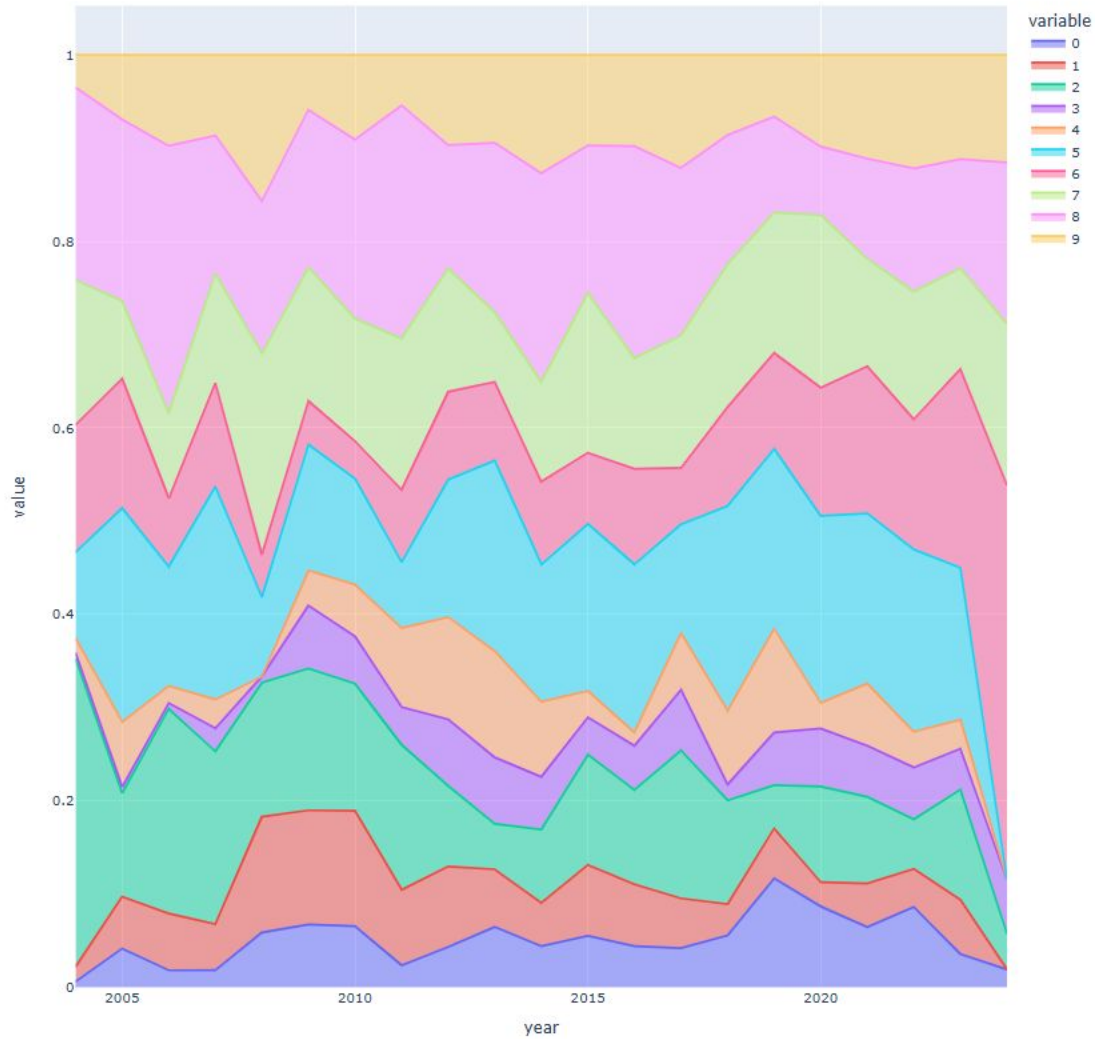
Categorization of actants by LLMs

o actant_name	o actant_name_std	o category
ong	ong	Non-Governmental Organizations (NGOs) and Advocacy Groups
citoyens	citoyen	Stakeholders by Demographic or Interest Groups
écologistes	écologiste	Non-Governmental Organizations (NGOs) and Advocacy Groups
environnement	environnement	Health and Environmental Focus Areas
agriculteurs	agriculteur	Stakeholders by Demographic or Interest Groups
greenpeace	greenpeace	Non-Governmental Organizations (NGOs) and Advocacy Groups
changement climatique	changement climatique	Health and Environmental Focus Areas
scientifiques	scientifique	Professional Roles and Actors
environnementalistes	environnementaliste	Non-Governmental Organizations (NGOs) and Advocacy Groups
commission européenne	commission européenne	Governmental Bodies and Public Authorities
pesticides	pesticide	Health and Environmental Focus Areas

Mapping actants & arguments



Clusters of arguments by year



Detecting periods

