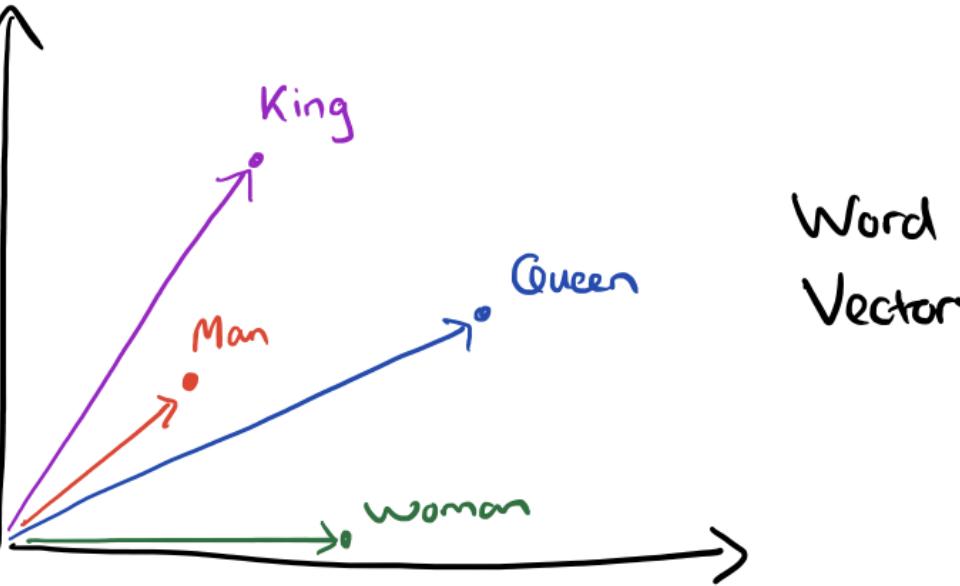
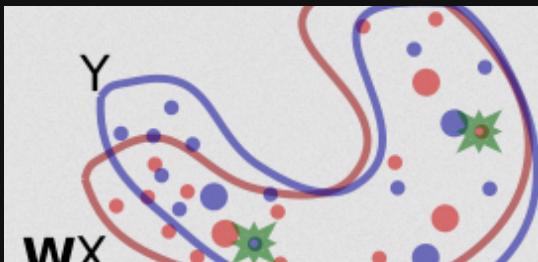


{ LLMs et Antiquité }

Utilisation des LLMs pour les lettres classiques



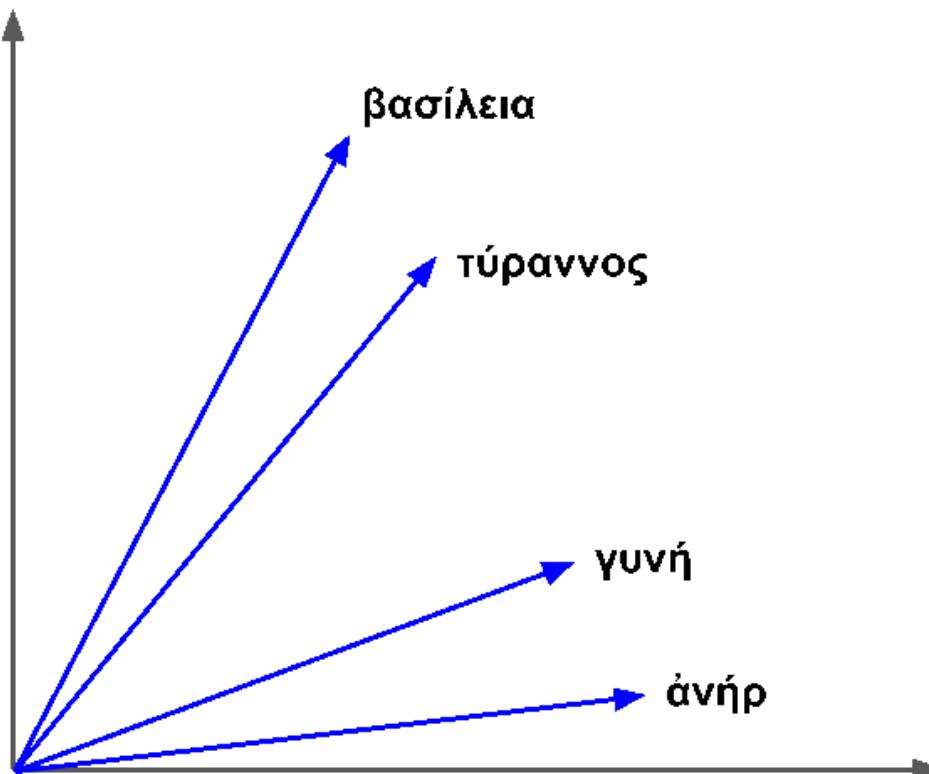
Word
Vector



Concepts de base

vecteurs statiques, vecteurs contextuels

Vecteurs statiques



Où vous trouverez tout

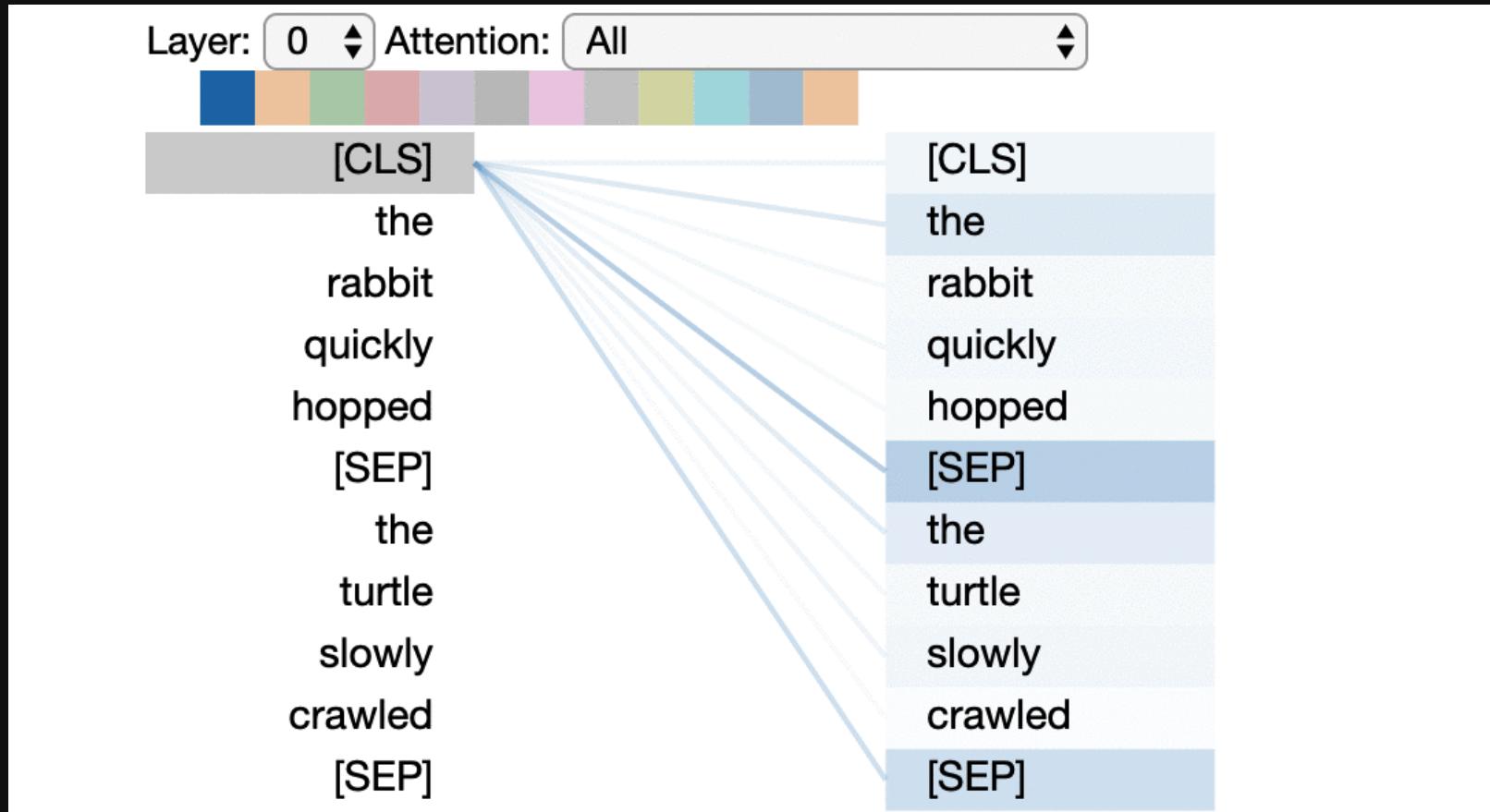
Vous trouverez tous les codes pour cet
atelier ici :

<https://tinyurl.com/cluster5b7>

Exemple avec un petit corpus (Homère)

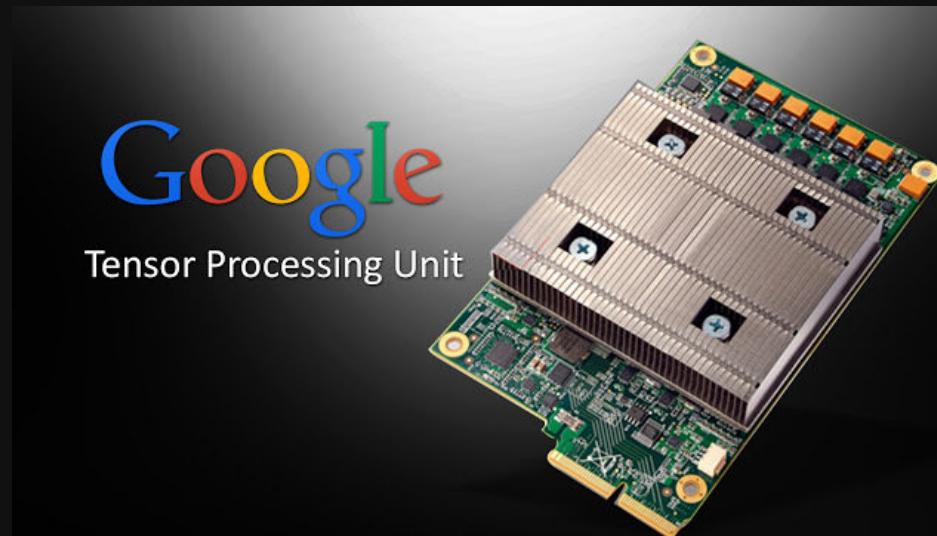
Il s'agit du notebook 1 dans le dépôt
disponible ici : <https://tinyurl.com/cluster5b7>

Définition d'un LLM (Large Language Model)



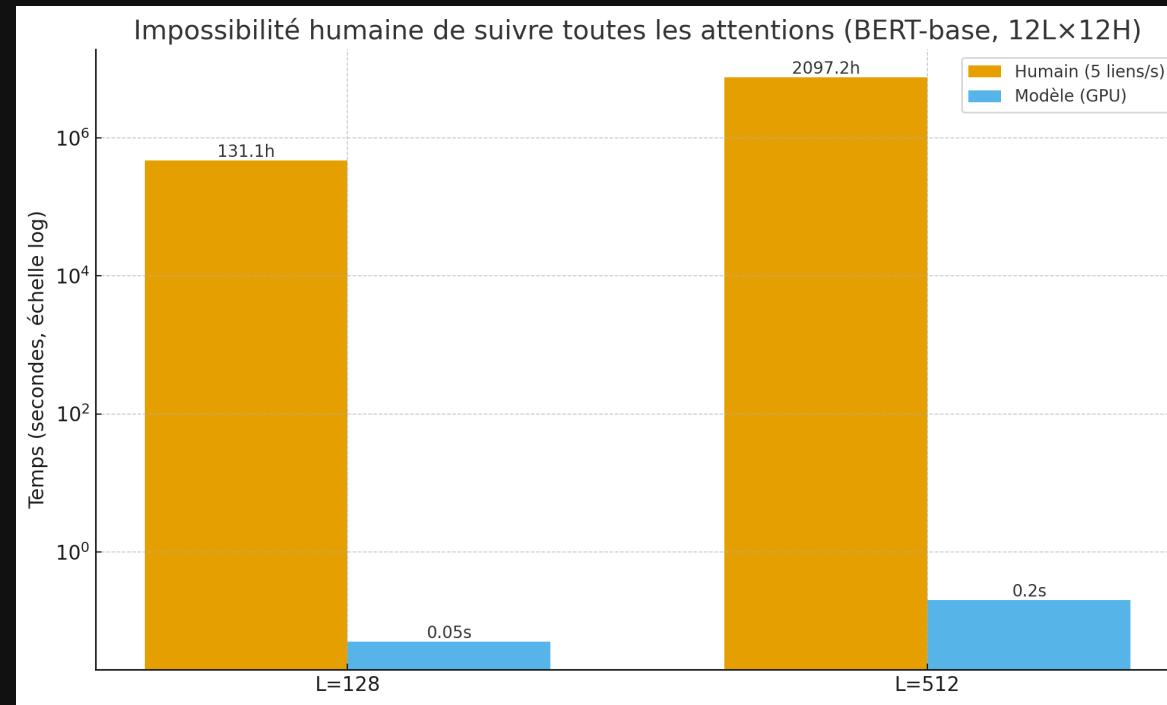
Problèmes ?

Vous aurez besoin d'un hardware digne de ce nom (et encore)



Problèmes ?

Il y a toujours un effet boîte noire dans les LLMS : même si on a de plus en plus de possibilités de visualiser les points d'attention décisionnels, physiquement, on ne peut pas tout contrôler.



Problèmes ?

En lettres classiques :

- très peu de données
- des données vraisemblablement plus finies que les autres
- trop peu de données représentatives (hors de la période classique)

	Class	Frequency	3 Classes	Frequency
ysssey	Person	2.469	PER	2
	Place	698	LOC	2
pnosophists	Person	12.424	PER	12
	Place	2.305	LOC	2
	Ethnic	3.548	MISC	6
	NoClass	2.263	MISC	6
	Group	681	MISC	6
	Title	206	MISC	6
	Festival	20	MISC	6
	Month	8	MISC	6
	Language	7	MISC	6
	Constellation	2	MISC	6
cal				24

Table 1: An overview of the training data set.

Ancient Greek BER1

Note: The Morphological Analysis Tag due to an issue with the FLAIR Toolkit help!



τελετέος	σε λαβών
Ἀχιλῆος	d ' Achille
Δαρδανίδαο ἥγε λαβών δ'	pris
έρινεὸν	à un figuier sauvage
φεύγοντι	qui s ' échappe
Λυκάονι	Lycaon
τόν	qu
ποτ'	naguère
αὐτὸς	lui - même
πατρός	de père
ἀλωῆς	verger

VM DE BELLO CIVILI LIBER PRIMVS

Word lookup Lookup

(Latin) Change Language

46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73

bis contentione, ut in senatu recitarentur; ut vero ex illi
publicae se non defuturum pollicetur, si audacter ac fortiter sententias dicere velint; sin Caesarem
capturum neque senatus auctoritati obtemperaturum: habere se quoque ad Caesaris gratiam
non deesse.

Incitat

incito , incitare , incitavi , incitatus (lesser)
verb; 1st conjugation
incito : enrage; urge on; inspire; arouse;
incit·ant
3rd person plur. pres. ind. act.

M. Marcellus, ingressus in urbem, decernere auderet, ut M. Iulium reservare et retinere vell. Lentulus sententiam Caium impetrare compulsi invenerat. Antonius, Q. Cassius, trecenti annos collaudatur.

Login to save your words to your wordlist. Log In

atque incitat. Multi undique ex eis ipsos et ipsum comitium tribunis, vocibus et concursu terrentur de his rebus eum doceant: sex

LLMs

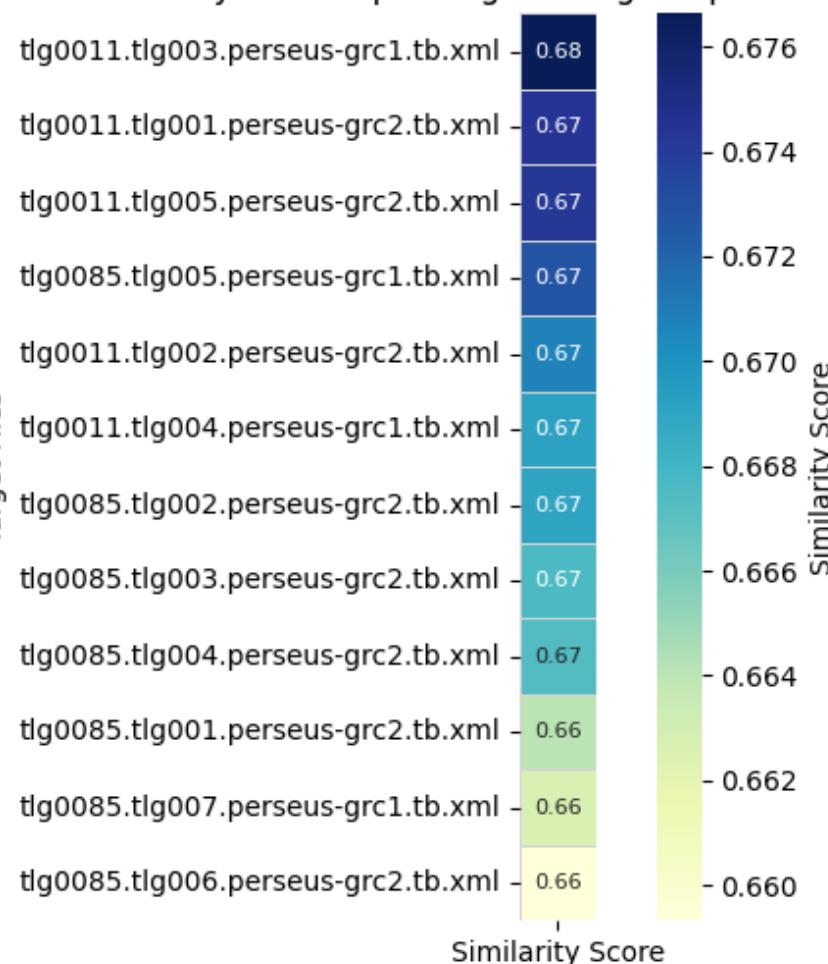
Quelques

exemples

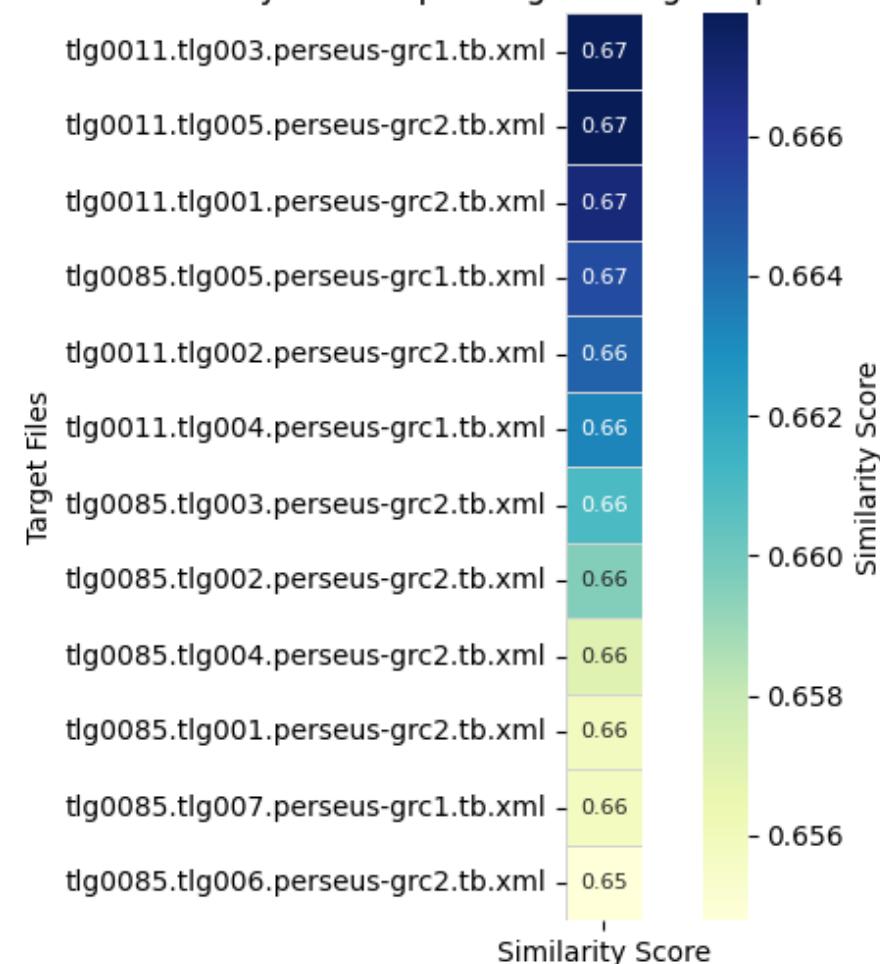
d'expériences

1. Similarité sémantique monolingue

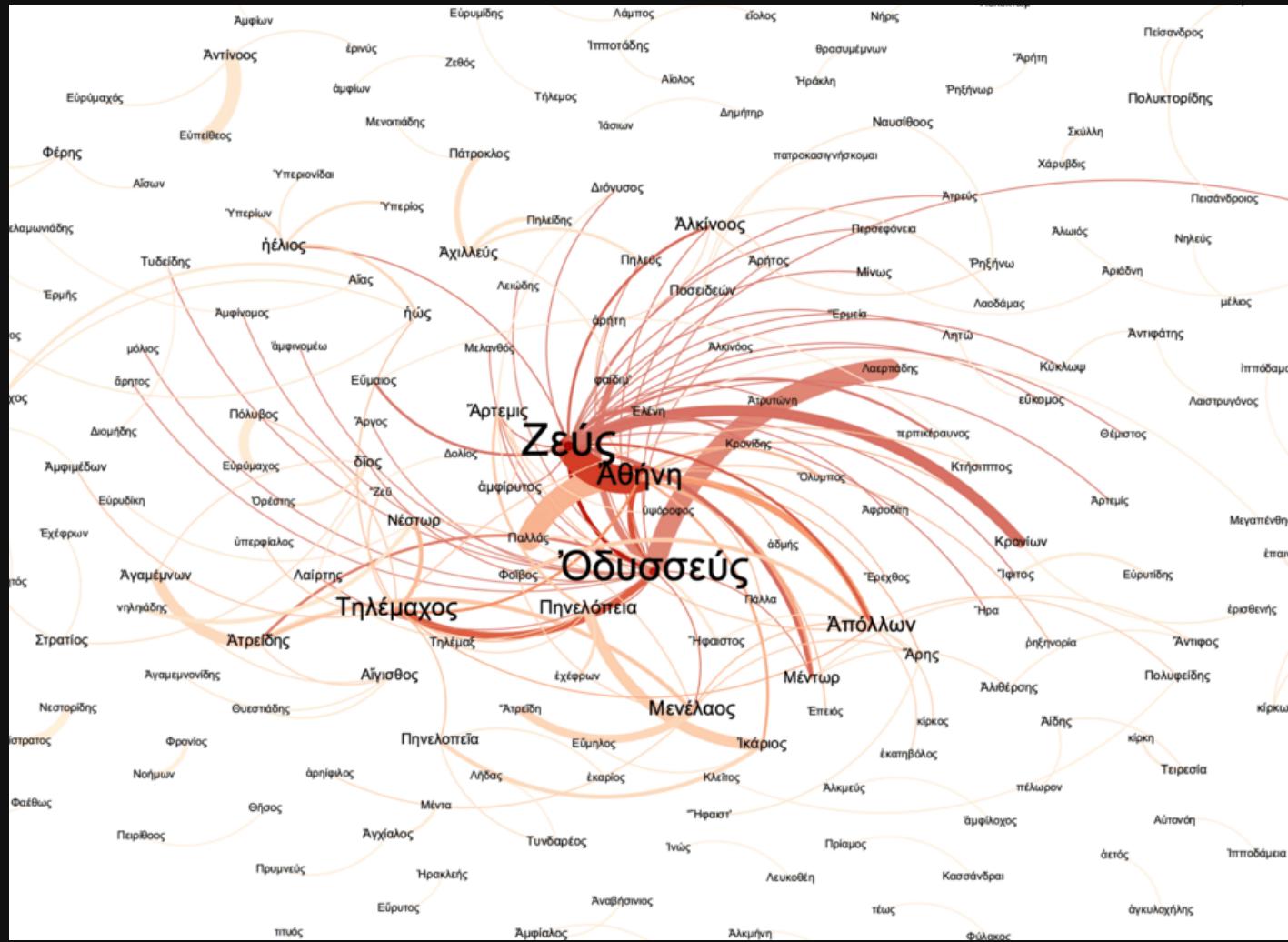
Similarity Heatmap for tlg0012.tlg001.perseus-grc1.tb.



Similarity Heatmap for tlg0012.tlg002.perseus-g



2. Etudes de réseaux

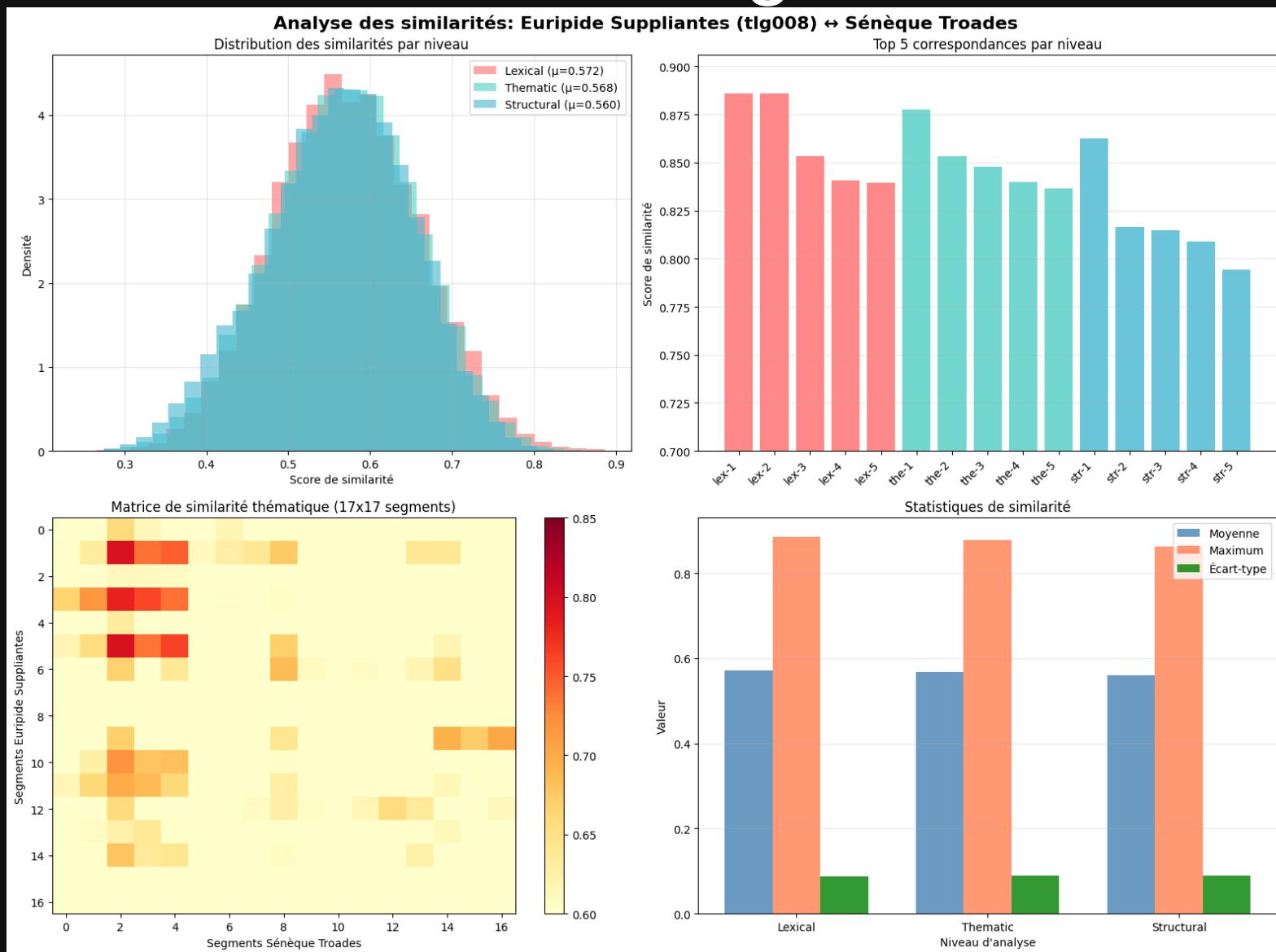


3. alignement de traductions

<https://github.com/OdysseusPolymetis/pyOdysseus>

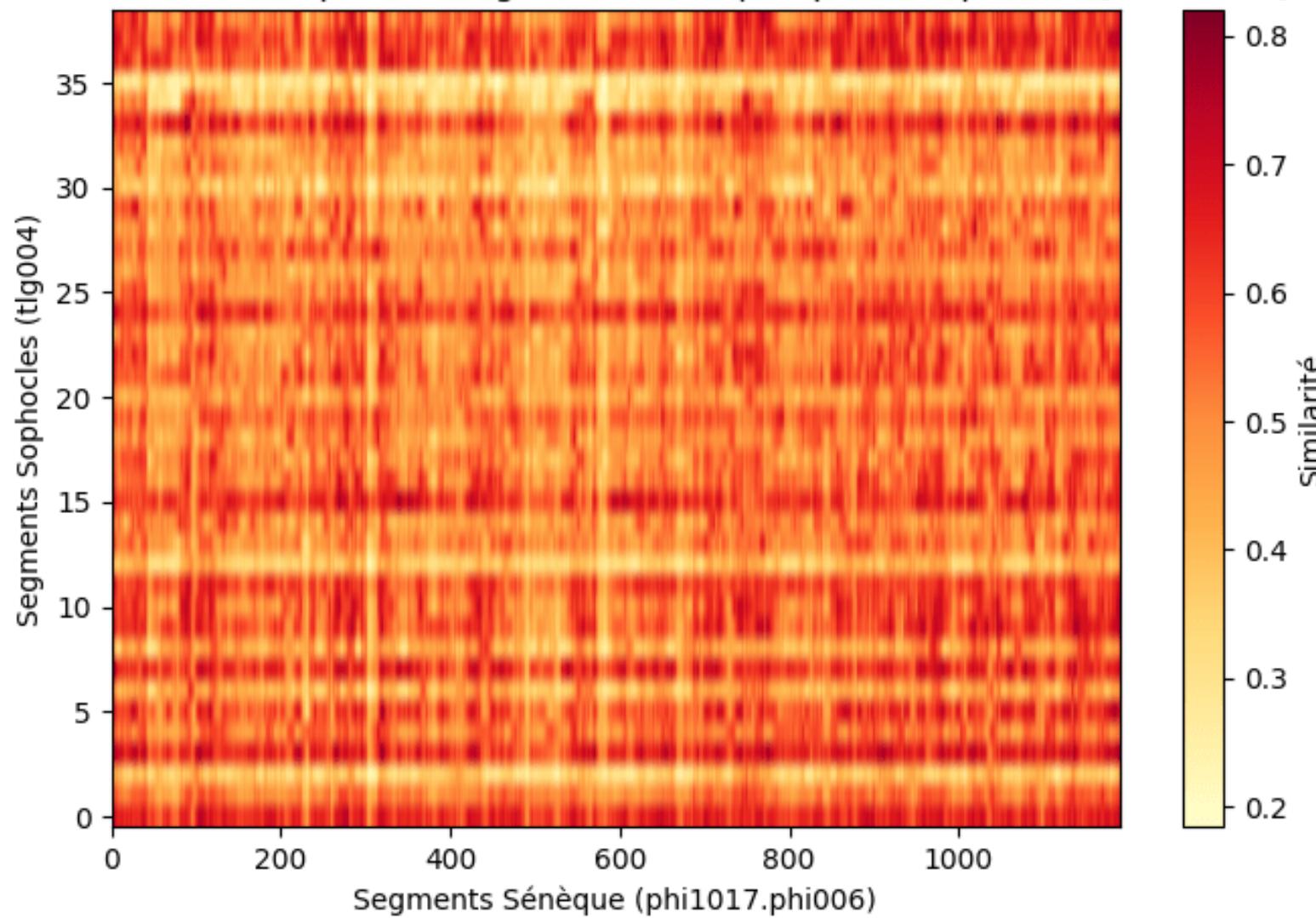
<https://odysseuspolymetis.github.io/pyOdysseus/paralogos/>

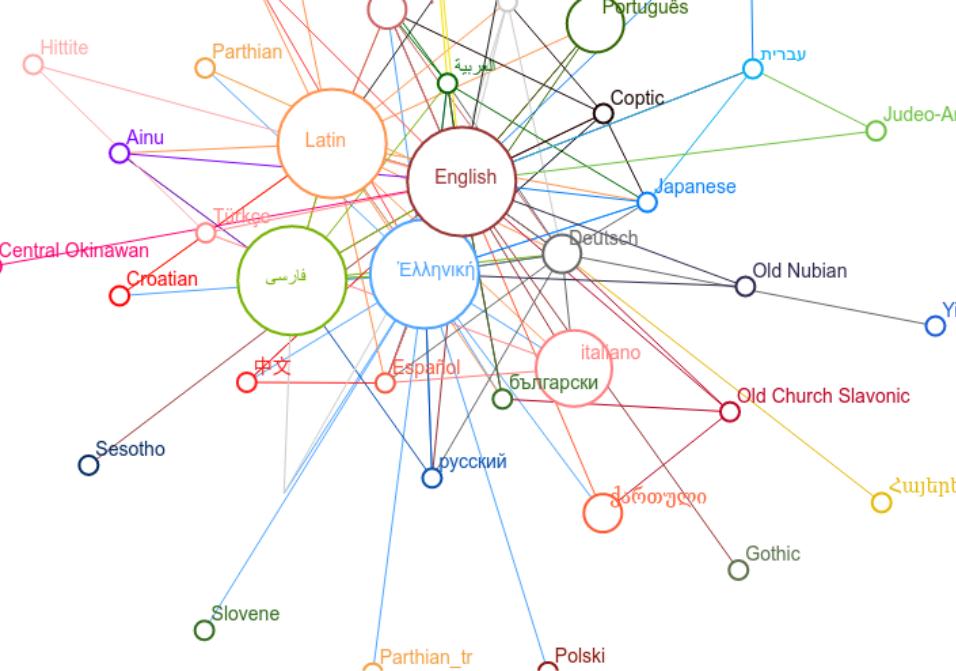
4. intertextualité multilingue



4. intertextualité multilingue

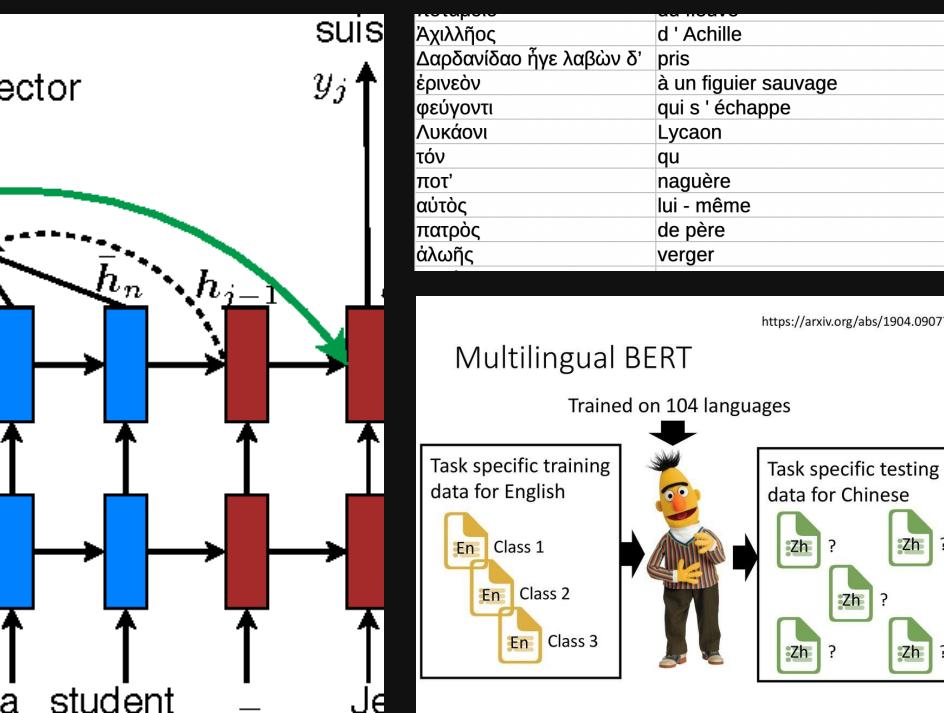
Carte de chaleur — Sophocles (tlg004) ↔ Sénèque (phi1017.phi006) [structural]





PRACTICE —

Atelier



Mise en pratique avec l'IA

Pour toutes les expériences qui vont suivre

Sur chacun des notebooks, vous avez un bouton en haut comme ça :



Vous devez appuyer dessus, ce qui ouvrira un Google Colab

(il vous faut un identifiant google)

Ensuite, n'oubliez pas de copier le notebook :
Fichier > Enregistrer une copie dans Drive

Expérience 1 : les vecteurs statiques

Vous trouverez tout sur le github, mais voici
le lien direct au notebook :

https://github.com/OdysseusPolymetis/journées_cluster5b_7/blob/main/1_word_vectors_from_treebank.ipynb

"A word is defined by the company it keeps"

J.R. Firth, linguiste britannique

Contexte des mots

Le **chat** mange des croquettes.

Le **chat** dort sur le canapé.

Le **chien** aboie contre le facteur.

Représentation vectorielle



Comment fonctionnent les word embeddings?

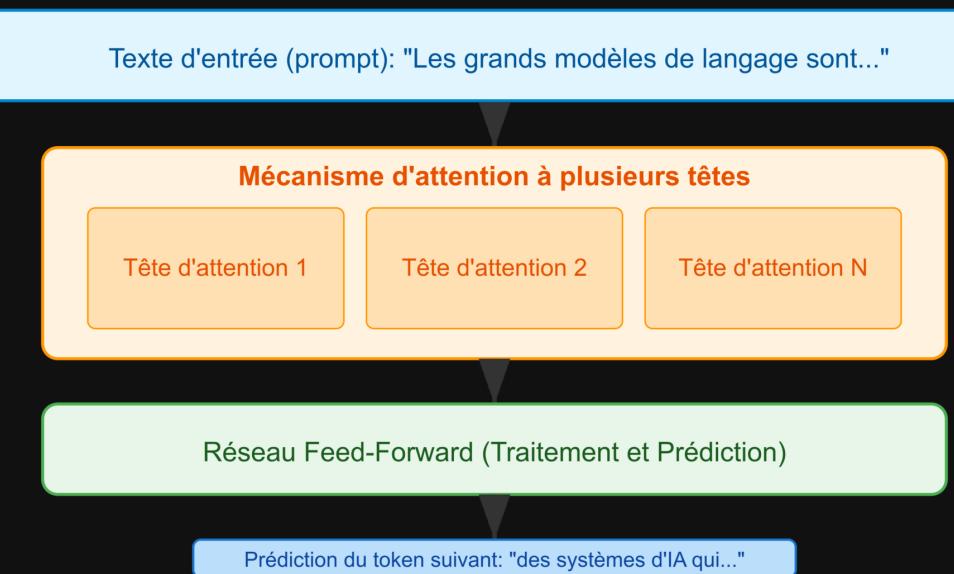
1. Chaque mot est représenté par un vecteur dans un espace à 100-300 dimensions
2. Les mots qui apparaissent dans des contextes similaires auront des vecteurs proches
3. Ces vecteurs capturent des relations sémantiques:

roi - homme + femme ≈ reine

chat = [0.2, -0.5, 0.8, 0.1, ...] (dans un espace à plusieurs dimensions)

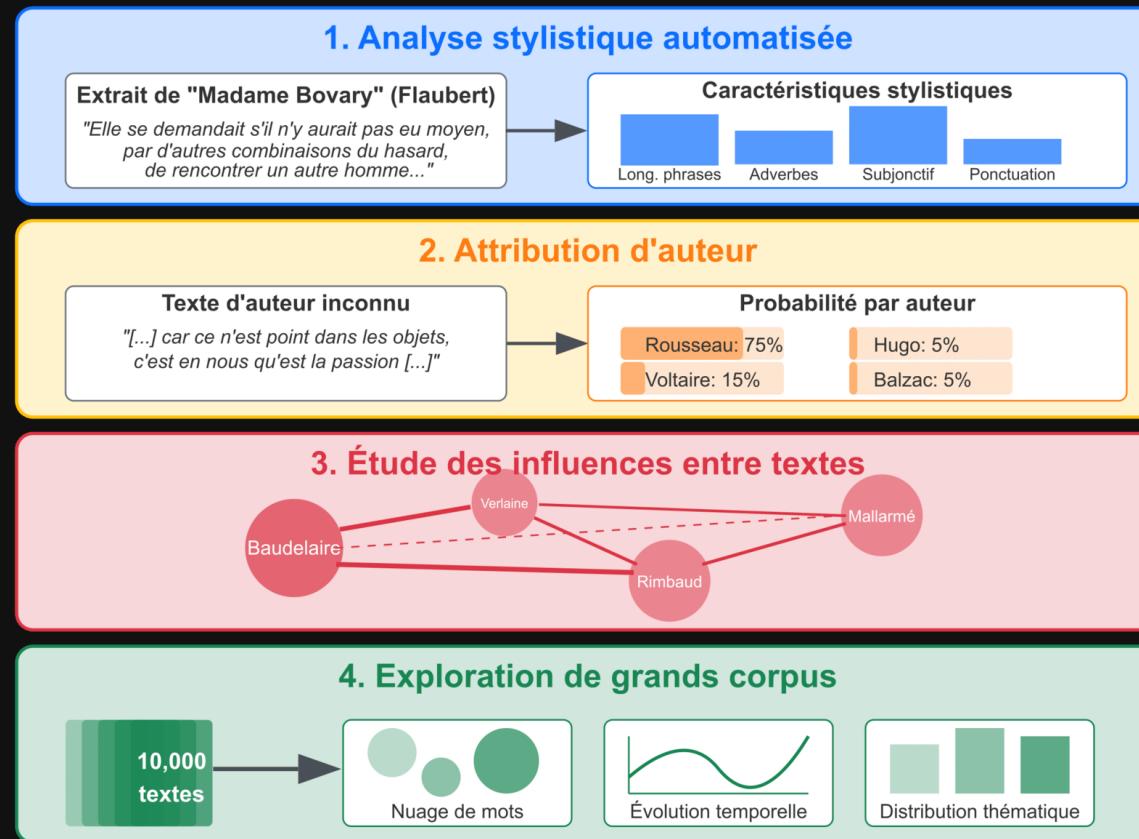
Expérience 2 : les transformers, intro

https://github.com/OdysseusPolymetis/journées_cluster5b_7/blob/main/2_bases_transformers.ipynb



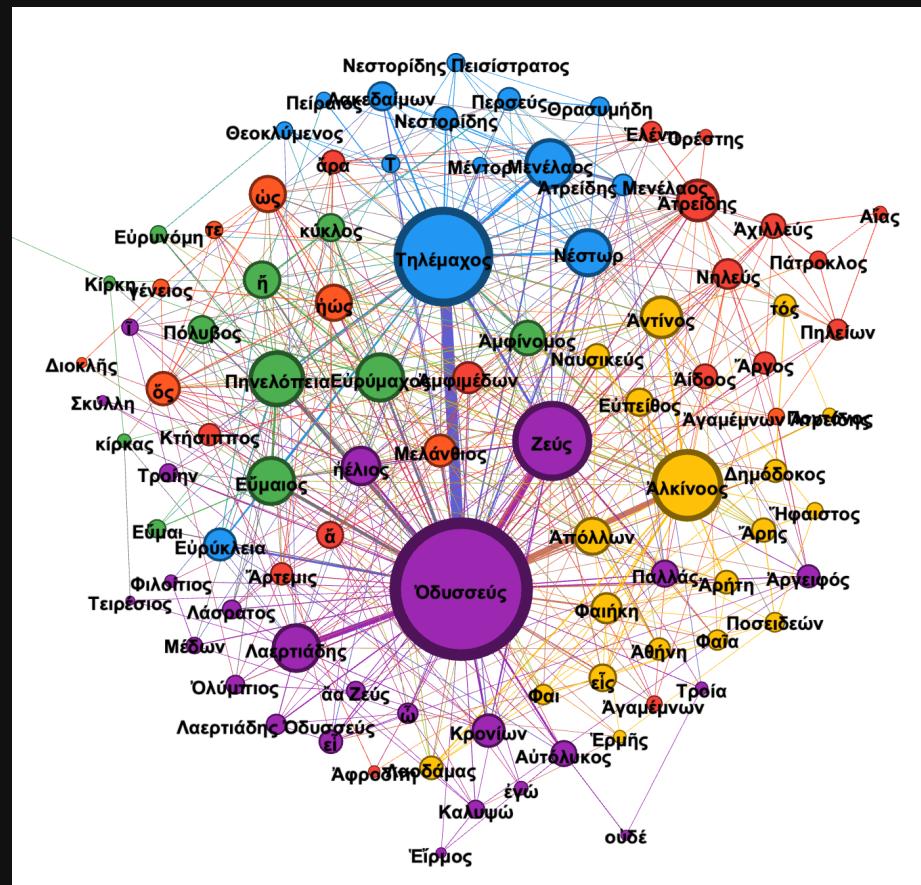
Expérience 3 : tal avec l'IA

https://github.com/OdysseusPolymetis/journées_cluster5b_7/blob/main/3_nlp_lat_gk.ipynb



Expérience 4 : NER

https://github.com/OdysseusPolymetis/journals_cluster5b_7/blob/main/4_ner_lat_gk.ipynb



Références utiles (non exhaustif)

- Yousef, T., Palladino, C., Shamsian, F., d'Orange Ferreira, A., & Ferreira dos Reis, M. (2022). An automatic model and Gold Standard for translation alignment of Ancient Greek. In Proceedings of the Language Resources and Evaluation Conference (pp. 5894–5905). European Language Resources Association. Marseille, France. Retrieved from <https://aclanthology.org/2022.lrec-1.634>
- Nils Reimers and Iryna Gurevych. 2020. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4512–4525, Online. Association for Computational Linguistics.
- Riemenschneider, F., & Frank, A. (2023). Graecia capta ferum victorem cepit. Detecting Latin Allusions to Ancient Greek Literature. In *Proceedings of the First Workshop on Ancient Language Processing*. Association for Computational Linguistics. Varna, Bulgaria. Retrieved from <https://arxiv.org/abs/2308.12008>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- MZES Data Lab. (n.d.). BERT and explainable AI. Retrieved from <https://www.mzes.uni-mannheim.de/socialsciencedatalab/article/bert-explainable-ai/>

Contact

Merci de votre attention

Colab link for experiment

marianne.reboul@ens-lyon.fr

[github : OdysseusPolymetis](#)