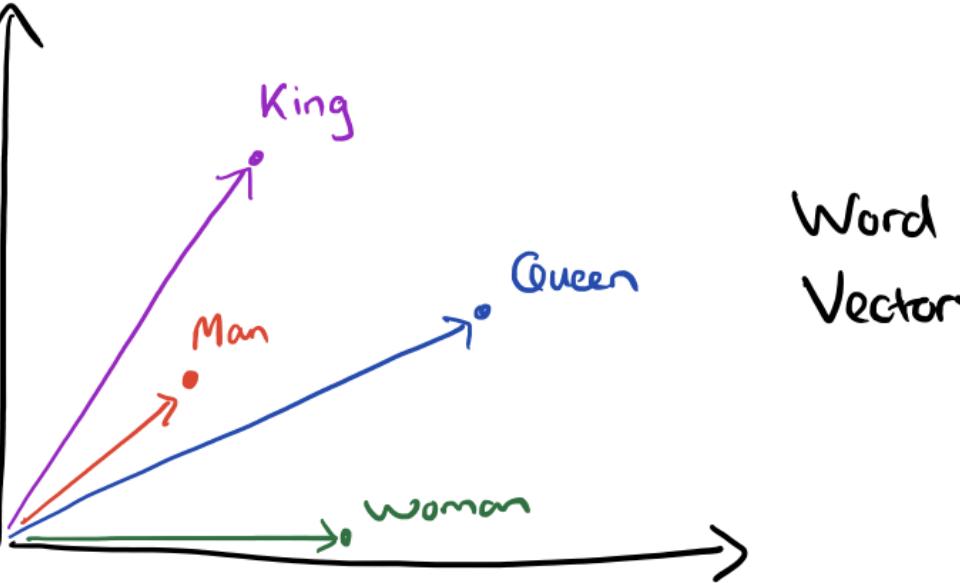
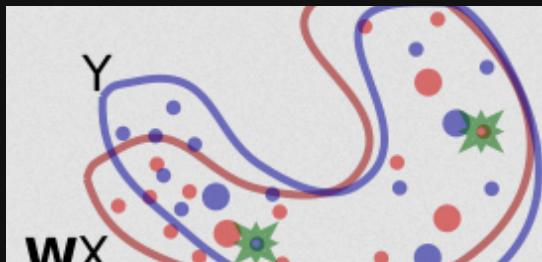


# { LLM, grec et latin }

*Constructions et usages de modèles multilingues pour le  
latin et le grec ancien*



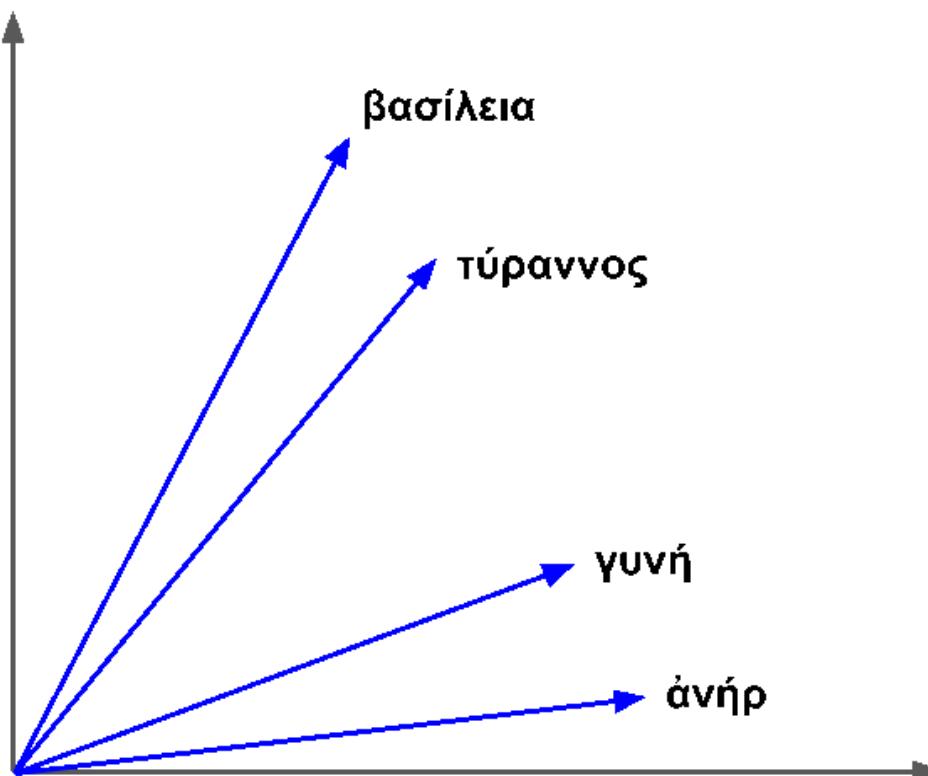
Word  
Vector



## Les bases

Vecteurs de mots statiques,  
contextuels...

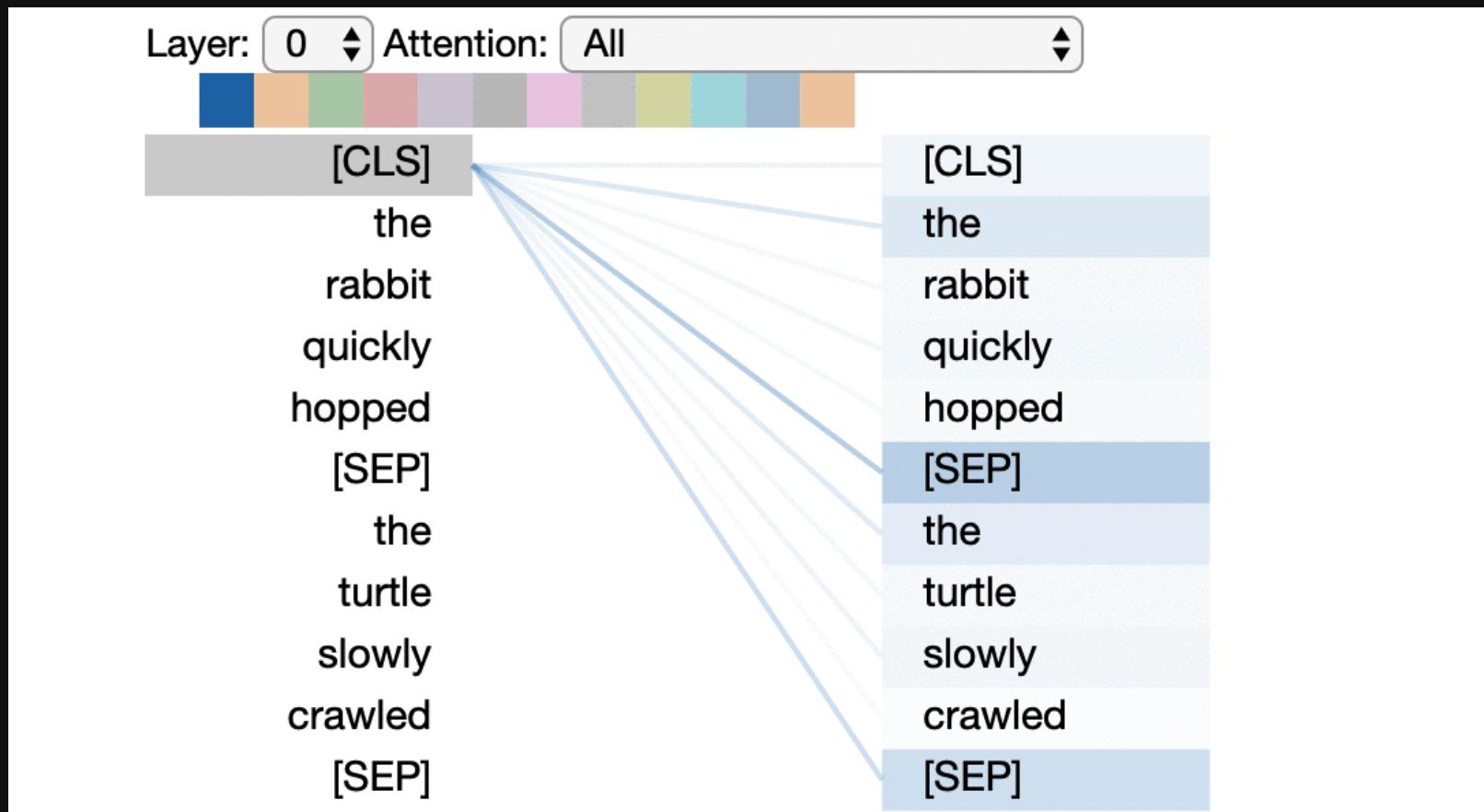
# Les vecteurs de mots



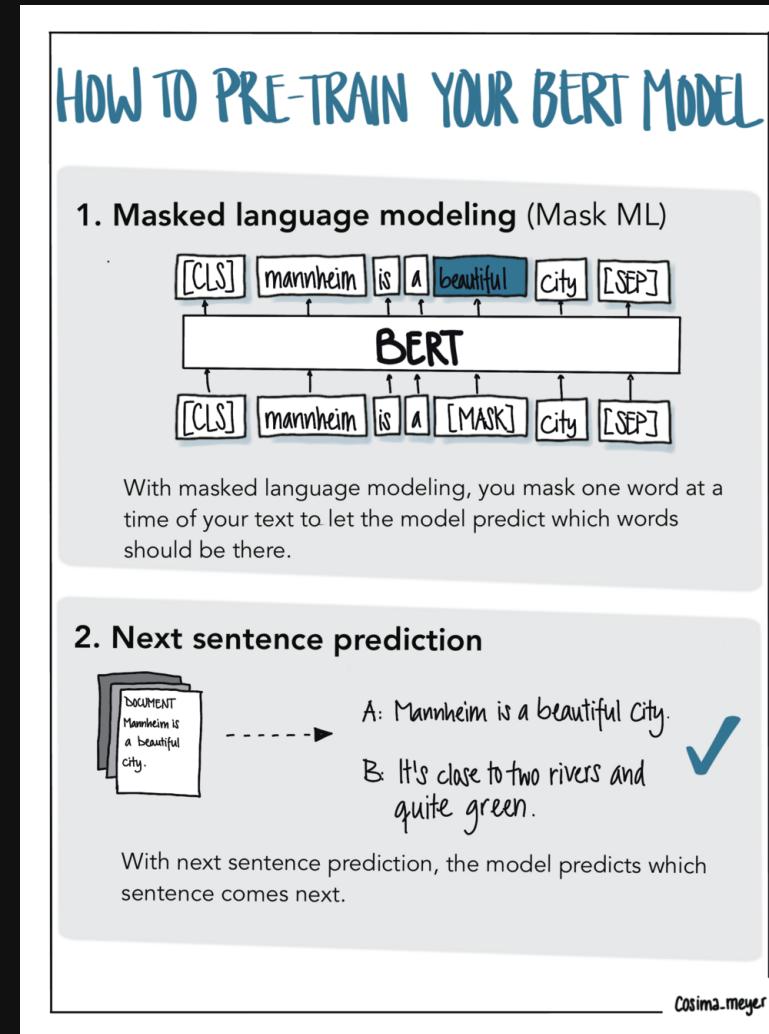
# Exemple sur un corpus conséquent

<https://projector.tensorflow.org/>

# Qu'est-ce qu'un grand modèle de langue (LLM) ?



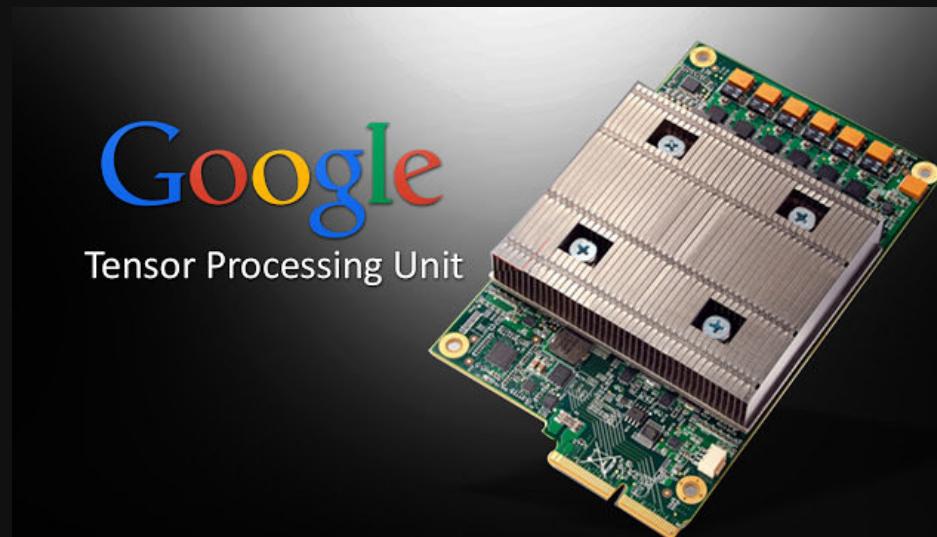
# Entraînement



<https://www.mzes.uni-mannheim.de/socialsciencedatalab/article/bert-explainable-ai/#bert>

# Problèmes ?

Petit problème matériel d'abord : les LLMs,  
ça n'est pas pour tout le monde.



# Problèmes ?

Mais le problème principal :

- le très faible nombre de données
- le très faible nombre de données représentatives (hors âge 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 Task due to an issue with the FLAIR Toolkit help!



τελευταῖς	σε μέσοις
Ἀχιλῆος	d' Achille
Δαρδανίδαιο ἥγε λαβών δ'	pris
έρινεόν	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

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

Incitare

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. Icilius reservare et retinere vell. Lentulus sententiam Caium impie plerique compulsi invenerunt. Antonius, Q. Cassius, trecenti viri collaudatur.

Log in to save your words to your wordlist. Log In

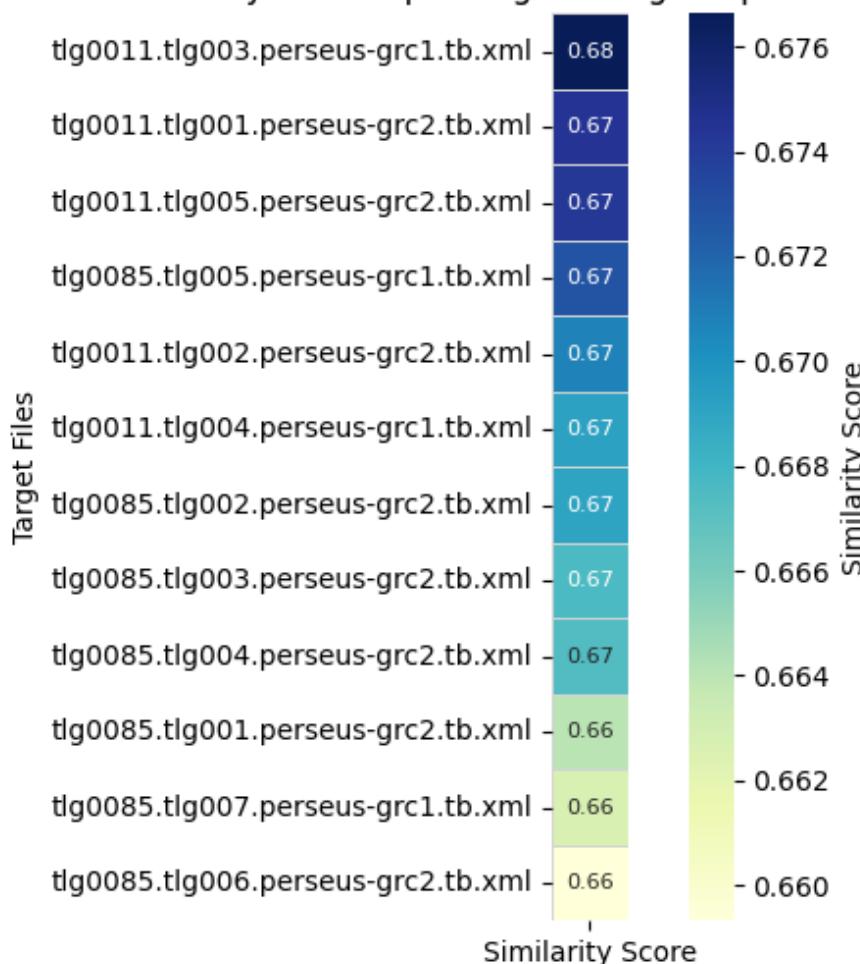
atque incitare. Multi undique ex eis et ipsum comitium tribunis, vocibus et concursu terrentur de his rebus eum doceant: sex

# Utilisation de LLMs monolingues

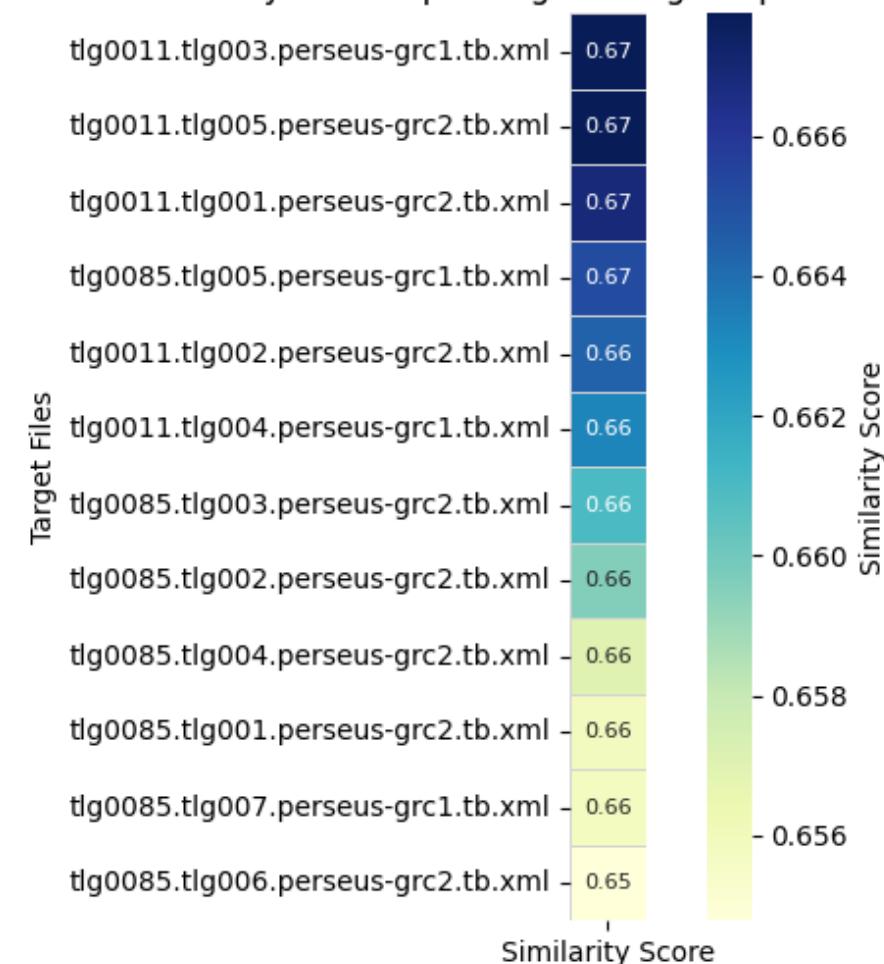
## À quoi cela peut servir ?

# Exemple 1 : similarité sémantique

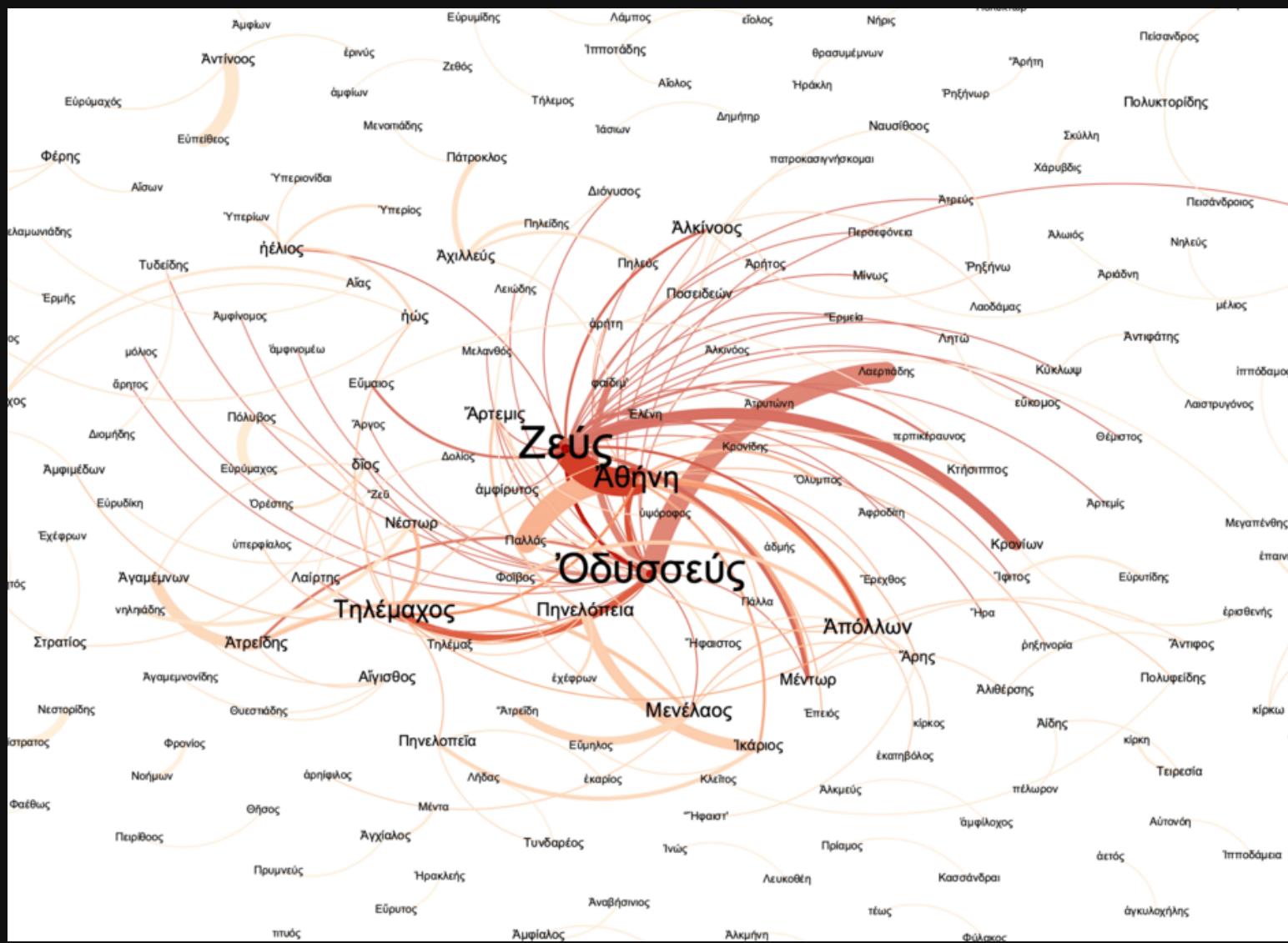
Similarity Heatmap for tlg0012.tlg001.perseus-grc1.tb.

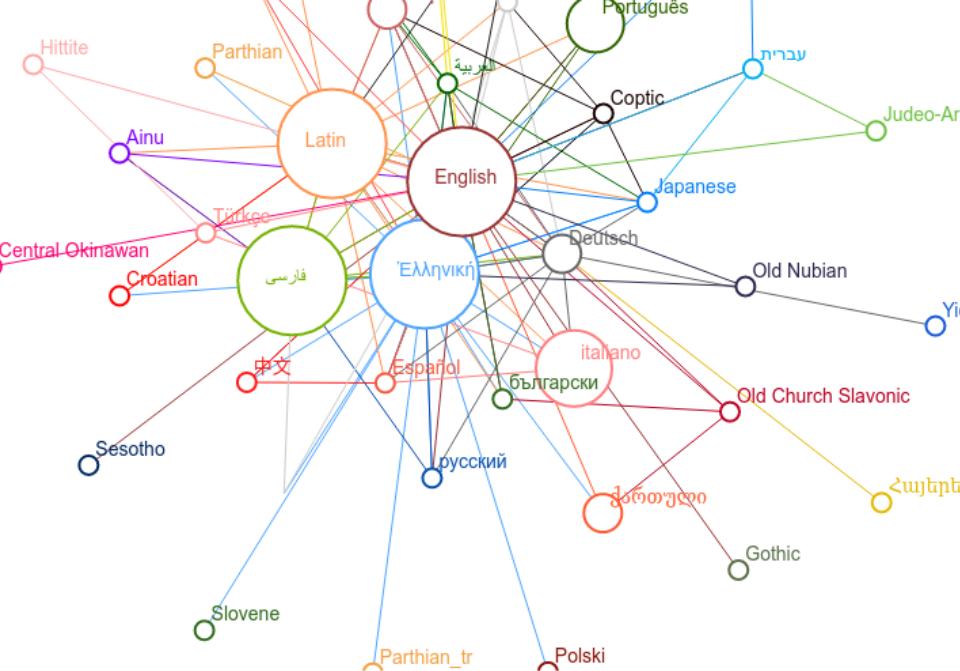


Similarity Heatmap for tlg0012.tlg002.perseus-g

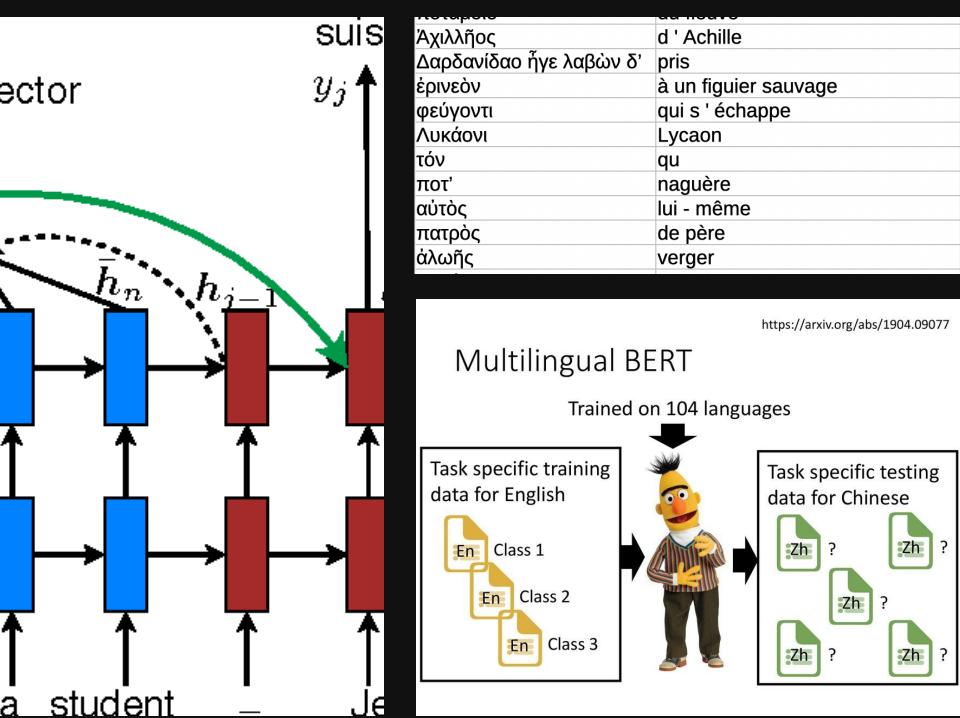


## Exemple 2 : NER



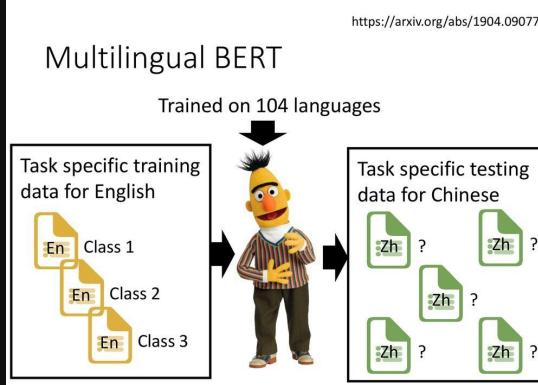


# MONOLINGUE



# Utilisation de LLMs multilingues

D'encore plus gros problèmes



# Pourquoi c'est encore plus compliqué

wmt19\_translate/de-fr

- **Description** de la configuration : ensemble de données de tâche de traduction de-en WMT 2019.
- **Taille du téléchargement** : 9.71 GiB

# Similarité sémantique au token

```
[ ] greek_word = "λόγος"
greek_embedding = greek_filtered_embeddings[greek_word] # Assurez-vous que cela est bien indexé

top_latin_words_with_scores = find_most_similar(greek_embedding, latin_filtered_embeddings)

for word, score in top_latin_words_with_scores:
    print(f"{word}: {score:.4f}")

oratio: 0.9894
τῷ: 0.9892
narratio: 0.9885
καὶ: 0.9878
liber: 0.9877
finis: 0.9877
scribo: 0.9877
loquor: 0.9875
ratio: 0.9875
argumentum: 0.9872
```

# Détection d'intertexte

Greek File: tlg0086.tlg025.perseus-grc1.xml, Latin File: phi0550.phi001.perseus-lat1.xml

Greek Phrase: δῆλον δ' ἔσται τὸ λεγόμενον ἐκ τῶν ὕστερον μᾶλλον.

Latin Phrase: id licet hinc quamvis hebeti cognoscere corde.

Similarity Score: 0.9741785526275635

Greek Phrase: δῆλον δ' ἔσται τὸ λεγόμενον ἐκ τῶν ὕστερον μᾶλλον.

Latin Phrase: id licet hinc quamvis hebeti cognoscere corde.

Similarity Score: 0.9741785526275635

Greek Phrase: περὶ μὲν οὖν τούτων δεδήλωται καὶ πρότερον·

Latin Phrase: id quod iam supra tibi paulo ostendimus ante.

Similarity Score: 0.9737507104873657

Greek Phrase: δῆλον δ' ἔσται τὸ λεγόμενον ἐκ τῶν ὕστερον μᾶλλον.

Latin Phrase: quae tibi posterius largo sermone probabo.

Similarity Score: 0.9700868129730225

Greek Phrase: περὶ μὲν οὖν τούτων δεδήλωται καὶ πρότερον·

Latin Phrase: id quod iam supera tibi saepe ostendimus ante.

Similarity Score: 0.9692978858947754

Greek Phrase: περὶ μὲν οὖν τούτων δεδήλωται καὶ πρότερον·

Latin Phrase: id quod iam supera tibi paulo ostendimus ante.

Similarity Score: 0.9681644439697266

# Contact

Je vous remercie.

<https://github.com/OdysseusPolymetis/ganes>  
hs\_ia\_02\_04\_24

marianne.reboul@ens-lyon.fr

github : OdysseusPolymetis