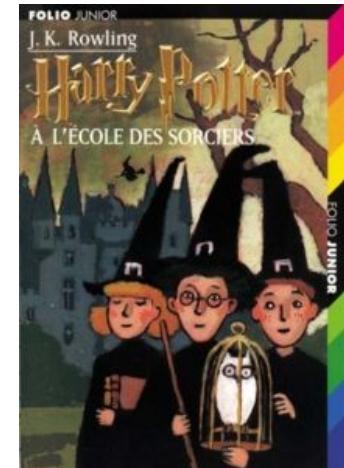


Projet HDDL Context Cemantix

Wassim Fathi - Arthur Augusto - Mael Montillet

I - Motivations

Motivation du projet



Motivation du projet



Livre utilisé : *Warbreaker* (2012), Brandon Sanderson

→ Roman de fantasy en libre accès

→ 238 000 mots, en anglais

Modélisation du jeu

Objectif : • Avoir un espace latent des mots où la proximité vectorielle dépend de l'œuvre

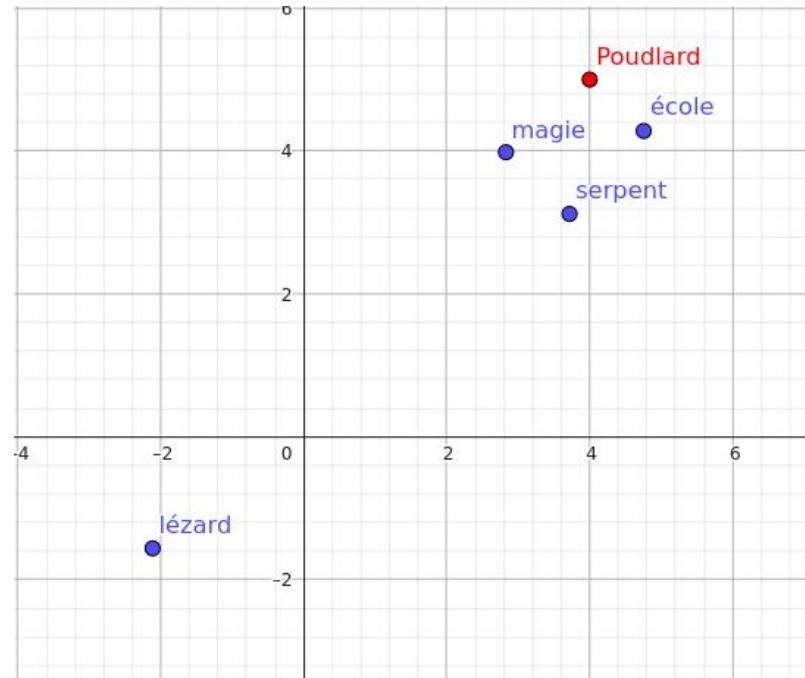
Métrique :

- Cosine Similarity : $\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$ \longleftrightarrow °C du Cémantix

$|\cos(\theta)|$ proche de 1 → haute température
 $|\cos(\theta)|$ proche de 0 → faible température



Modélisation du jeu



II - Choix du modèle

Présentation de l'architecture

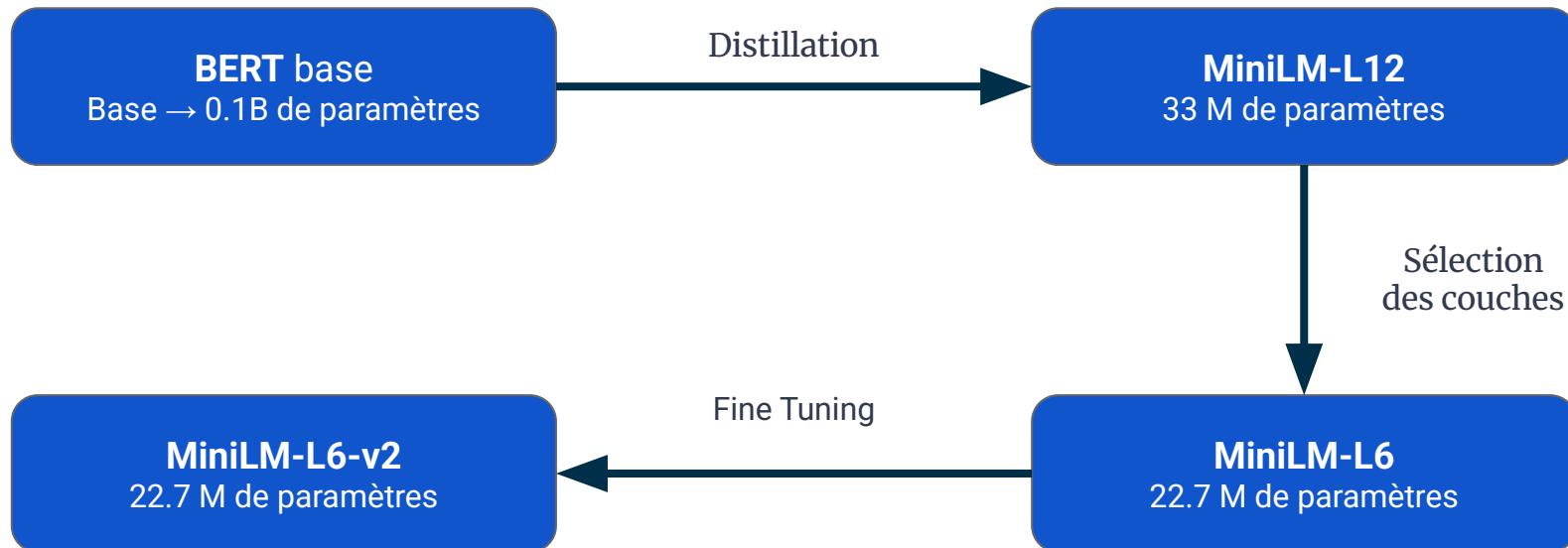
Word Embedding Models

Option A : Word2Vec (vocabulaire fini)

Option B : FastText (peu de profondeur contextuelle)

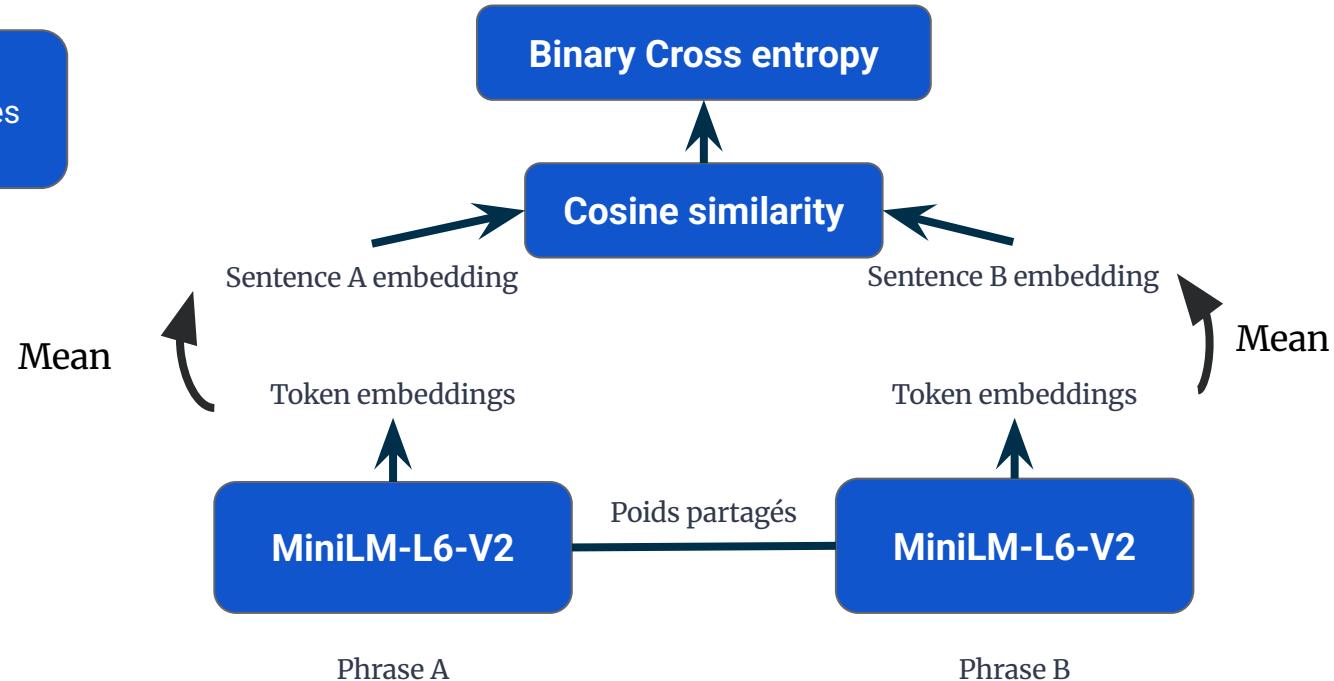
Option C : BERT / MiniLM (Etat de l'art sur la précision sémantique)

Modèle pré-entraîné



Siamese network fine-tuning

Dataset : 1B paires de phrases



Sentence transformers

Sentence transformers : utiliser des structures de réseau siamois ou triplet pour créer des représentations de phrases sous forme de vecteurs sémantiquement significatifs qui peuvent être comparés à l'aide de la similarité cosinus.

III - Entraînement du modèle

Datasets d'entraînement

warbreaker.txt



27 000 phrases dans l'ordre :

- 1: Why, Vasher thought, do so many things begin with me getting thrown into prison?
- 2: The guardsmen laughed to one another outside, slamming the cell door shut with a clang.

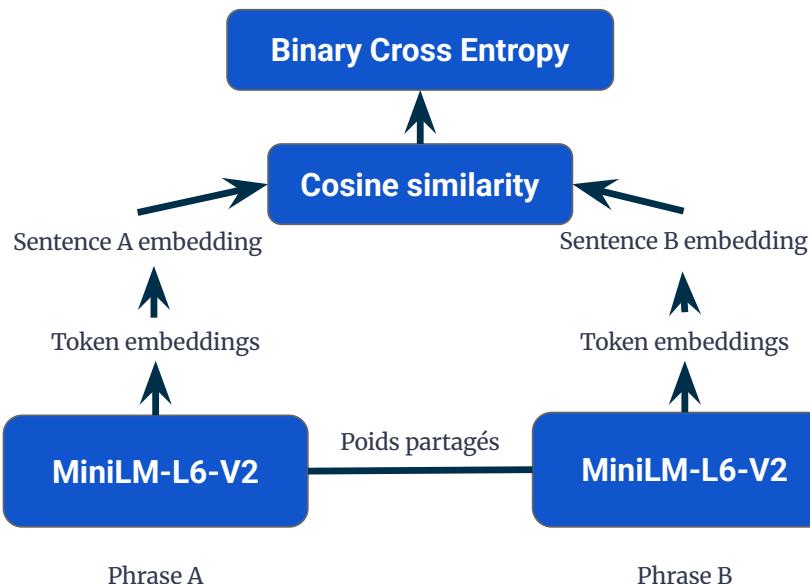
- 3: Vasher stood and dusted himself off, rolling his shoulder and wincing.
- 4: He glanced about.

- 5: While the bottom half of his cell door was made of solid wood, the top half was barred.
- 6: One of them noticed him.

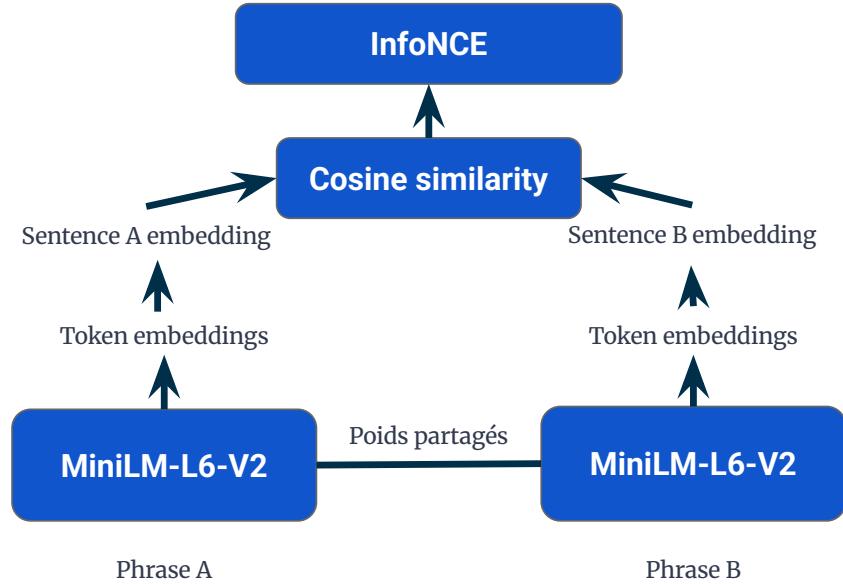
Paire positive
(2 phrases consécutives)

Paire négative

Nos méthodes de fine-tuning



Méthode 1



Méthode 2

Nos méthodes de fine-tuning

Méthode 1 - BCE loss

$$\mathcal{L}_{\text{BCE}}(y, \hat{y}) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

$y = 1 \rightarrow$ Paire positive

$y = 0 \rightarrow$ Paire négative

$$\hat{y} = \frac{1 + \text{sim}(z_1, z_2)}{2}$$

Inconvénients :

- Les paires sont traitées séparément
- Sensible à l'équilibrage positif / négatif

Nos méthodes de fine-tuning

Méthode 2 - InfoNCE loss

$$\text{loss}_i = - \log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_i^+)/\tau)}{\sum_{j=1}^{2N} \mathbf{1}_{[j \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}$$

Paire positive

Paire négative

Avantage :

- Met en compétition les paires négatives avec les positives
- Structure bien l'espace latent

Résultats

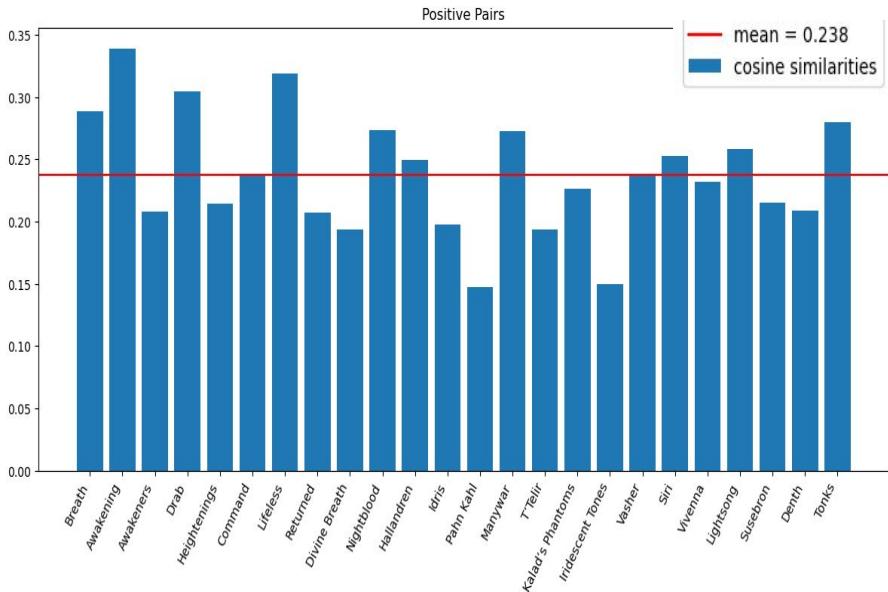
Méthodologie :

- Identification de mots spécifiques à l'univers de *Warbreaker* (*ex: Breath, Vasher*) → **24 mots**
- Création de paires positives (111) et négatives (80) pour chaque mot
Ex : ('*Breath*', '*Soul*'), ('*Breath*', '*Metal*')
- Comparaison de la cos similarité des paires positives/négatives avant/après entraînement du modèle

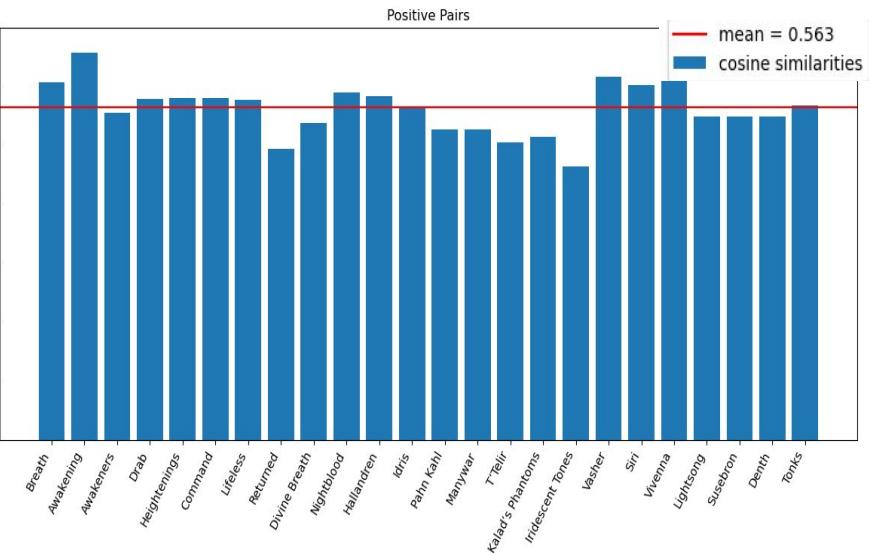
Résultats

Méthode 2 (Contrastive loss)

Avant fine-tuning (ancien modèle)



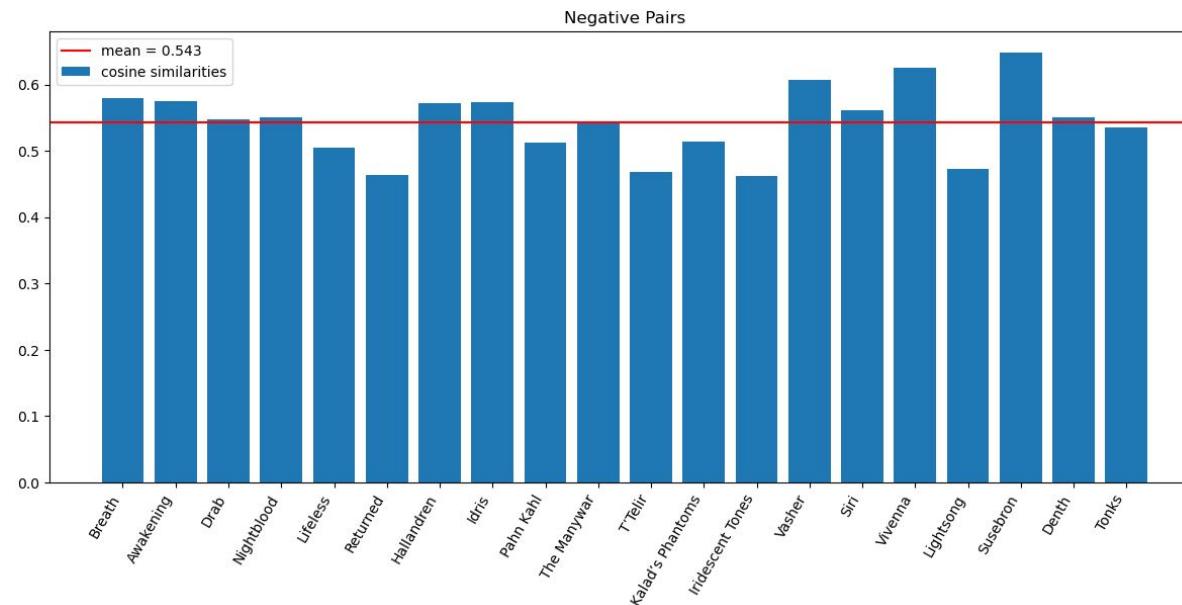
Après fine-tuning (modèle ré-entraîné)



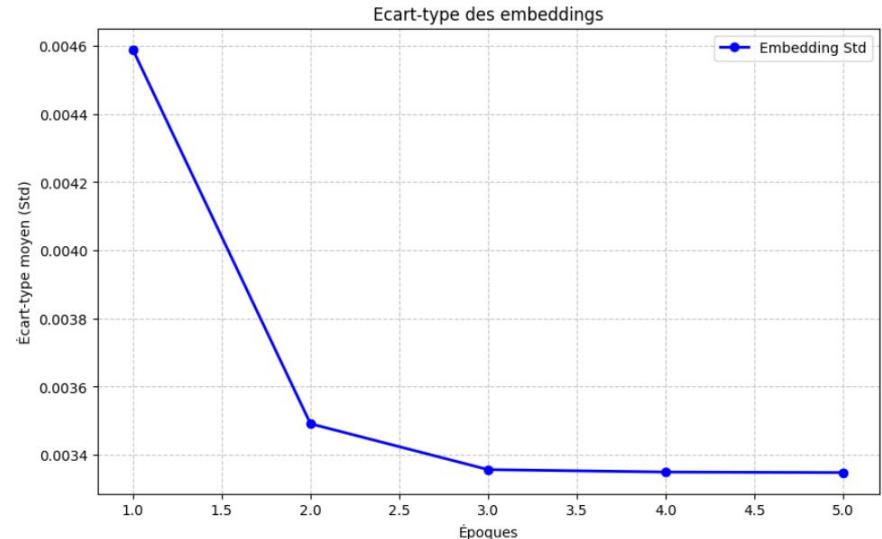
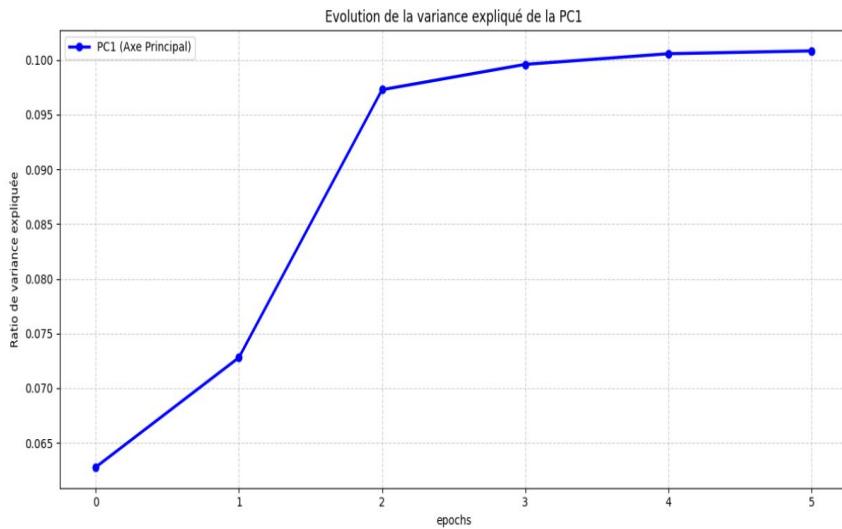
Résultats

Méthode 2 (Contrastive loss)

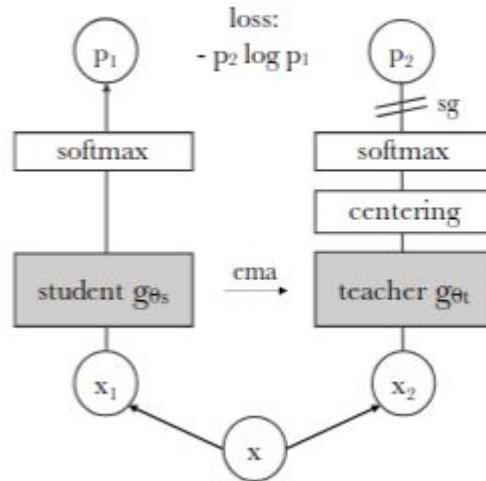
Après fine-tuning (modèle ré-entraîné)



Discussions



Perspectives



Dino v1 (positives pairs only)