

Présentation de nouveaux indicateurs d'écriture collaborative et du processus de leur construction



Atelier-Symposium EIAH 2023 « Les tableaux de bord de suivi en éducation »

Anis Mohamed HADDOUCHE **

Equipe projet :

- ▶ Fahima DJELLIL **
- ▶ Christian HOFFMANN *
- ▶ Nadine MANDRAN *
- ▶ Cédric HAM *

** IMT Atlantique , * Université Grenoble-Alpes

Table des matières

I. Introduction

II. Indicateurs d'écriture collaborative « non-sémantique » :
bref aperçu

III. Indicateur d'écriture collaborative « sémantique »

IV. Perspectives

I. Introduction

Contexte

- ▶ Nos travaux nous ont conduits à élaborer des indicateurs **qui peuvent être** utilisés pour détecter des **stratégies** d'écriture collaborative sur des documents textes
- ▶ Notre présentation se focalisera sur le processus de **construction** ces indicateurs

I. Introduction

Contexte

- ▶ En vue d'évaluer nos indicateurs, nous utilisons LabNbook, une plateforme numérique pour l'apprentissage des sciences expérimentales
- ▶ Chaque document (appelé **LabDoc**) qui ne peut être édité que par **un seul apprenant à la fois**

II. Indicateurs d'écriture collaborative « non-sémantique » : bref aperçu

Collaboration

- ▶ La combinaison de ces indicateurs permet de détecter deux stratégies de collaboration (Onrubia et Engel 2009)
 - ▶ Construction séquentielle **sommative** : un membre propose un document initial (complet ou pas), les autres ajoutent successivement leurs contributions **sans** modifier ce qui a été écrit auparavant
 - ▶ Construction séquentielle **intégrative** : un membre propose un document initial (complet ou pas), les autres ajoutent successivement leurs contributions **et** modifient ce qui a été écrit auparavant
- ▶ Nous construisons deux indicateurs pour **détecter** ces deux stratégies

II. Indicateurs d'écriture collaborative « non-sémantique » : bref aperçu

Comment mesurer la différence entre deux séquences ?

- ▶ distance de Levenshtein
- ▶ Librairie DiffLib de Python
- ▶ Recherche les plus grandes séquences similaires
- ▶ Accorder des tags pour chaque séquence

En **bleu**, on a indiqué les parties communes aux deux textes, en **rouge** les parties supprimées, en **vert** les parties insérées et en **orange** les parties modifiées.

Texte initial : « LabNbook est une plateforme gratuite, utilisée par plus de 2 800 **élèves** chaque année, à l'Université Grenoble-Alpes, Grenoble-INP, dans des **collèges**, lycées et des CPGE. »

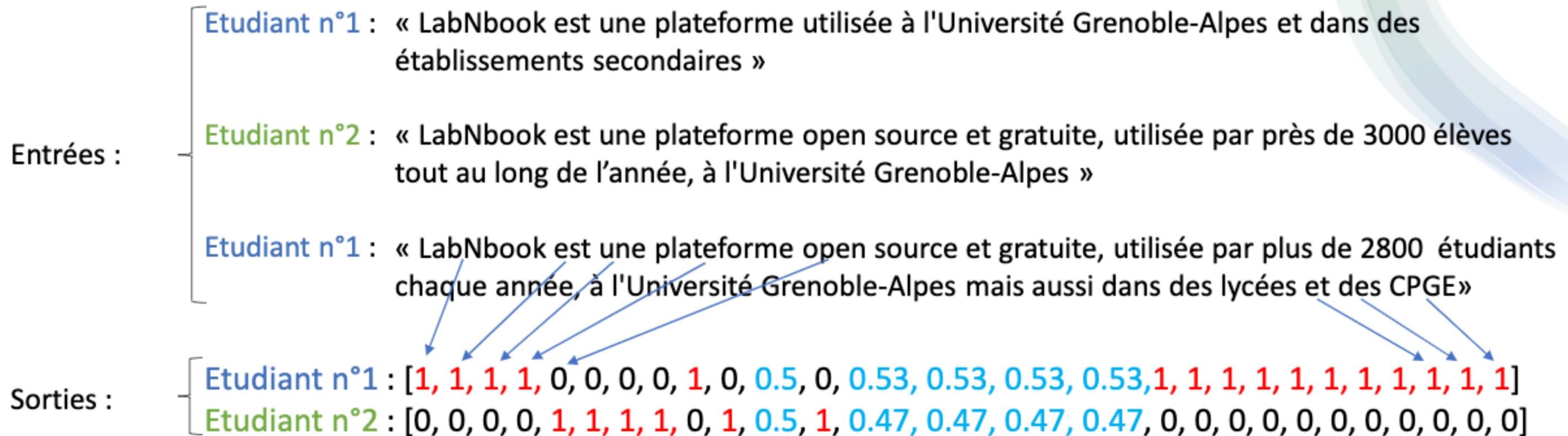
Texte final : « LabNbook est une plateforme **open source et** gratuite, utilisée par plus de 2 800 **étudiants** chaque année, à l'Université Grenoble-Alpes, Grenoble-INP, dans des lycées et des CPGE. »

equal	'LabNbook est une plateforme' --> 'LabNbook est une plateforme'
insert	" --> 'open source et'
equal	'gratuite, utilisée par plus de 2 800' --> 'gratuite, utilisée par plus de 2 800'
replace (0.73)	'élèves' --> 'étudiants'
equal	'chaque année, à l'Université ... dans des' --> 'chaque année, à l'Université ... dans des'
delete	'collèges,' --> ''
equal	'lycées et des CPGE.' --> 'lycées et des CPGE.'

II. Indicateurs d'écriture collaborative « non-sémantique » : bref aperçu

Comment agréger cette information pour la rendre exploitable ?

- ▶ Matrice de contribution :



II. Indicateurs d'écriture collaborative « non-sémantique » : bref aperçu

équilibre de contribution

- ▶ Mesure combien les contributions d'écriture collaborative des apprenants sont égales (bien équilibrées) ou non en termes de mots constituant le texte

- ▶ Matrice de contribution X

$$\begin{bmatrix} x_{1,1,1} & \dots & x_{1,1,l} & \dots & x_{1,1,n_1} & \dots & x_{1,j,1} & \dots & x_{1,j,l} & \dots & x_{1,j,n_j} & \dots & x_{1,N,1} & \dots & x_{1,N,l} & \dots & x_{1,N,n_N} \\ \vdots & \ddots & \vdots & \dots & \vdots & \dots & \vdots & \ddots & \vdots & \dots & \vdots & \dots & \vdots & \ddots & \vdots & \dots & \vdots \\ x_{i,1,1} & \dots & x_{i,1,l} & \dots & x_{i,1,n_1} & \dots & x_{i,j,1} & \dots & x_{i,j,l} & \dots & x_{i,j,n_j} & \dots & x_{i,N,1} & \dots & x_{i,N,l} & \dots & x_{i,N,n_N} \\ \vdots & \dots & \vdots & \ddots & \vdots & \dots & \vdots & \dots & \vdots & \ddots & \vdots & \dots & \vdots & \dots & \vdots & \ddots & \vdots \\ x_{K,1,1} & \dots & x_{K,1,l} & \dots & x_{i,1,n_1} & \dots & x_{K,j,1} & \dots & x_{K,j,l} & \dots & x_{i,j,n_j} & \dots & x_{K,N,1} & \dots & x_{K,N,l} & \dots & x_{K,N,n_N} \end{bmatrix}$$

- ▶ Équilibre de contribution $e(X)$

$$e(X) = 1 - \frac{K}{K-1} \sum_{i=1}^K \left(\bar{x}_{i,\cdot} - \frac{1}{K} \right)^2 \quad \text{où} \quad \bar{x}_{i,\cdot} = \frac{1}{N} \sum_{j=1}^N \sum_{l=1}^{n_j} x_{i,j,l}$$

II. Indicateurs d'écriture collaborative « non-sémantique » : bref aperçu

co-écriture

- ▶ l'indicateur de co-écriture mesure les contributions des auteurs au niveau des phrases.
- ▶ La construction de cet indicateur nécessite en amont le découpage du texte en phrases.

II. Indicateurs d'écriture collaborative « non-sémantique » : bref aperçu

Nos indicateurs : co-écriture

Mat. de contribution

$$\begin{bmatrix} x_{1,1,1} & \dots & x_{1,1,l} & \dots & x_{1,1,n_1} & \dots & x_{1,j,1} & \dots & x_{1,j,l} & \dots & x_{1,j,n_j} & \dots & x_{1,N,1} & \dots & x_{1,N,l} & \dots & x_{1,N,n_N} \\ \vdots & \ddots & \vdots & \dots & \vdots & \dots & \vdots & \ddots & \vdots & \dots & \vdots & \dots & \vdots & \ddots & \vdots & \dots & \vdots \\ x_{i,1,1} & \dots & x_{i,1,l} & \dots & x_{i,1,n_1} & \dots & x_{i,j,1} & \dots & x_{i,j,l} & \dots & x_{i,j,n_j} & \dots & x_{i,N,1} & \dots & x_{i,N,l} & \dots & x_{i,N,n_N} \\ \vdots & \dots & \vdots & \ddots & \vdots & \dots & \vdots & \dots & \vdots & \ddots & \vdots & \dots & \vdots & \dots & \vdots & \ddots & \vdots \\ x_{K,1,1} & \dots & x_{K,1,l} & \dots & x_{K,1,n_1} & \dots & x_{K,j,1} & \dots & x_{K,j,l} & \dots & x_{K,j,n_j} & \dots & x_{K,N,1} & \dots & x_{K,N,l} & \dots & x_{K,N,n_N} \end{bmatrix}$$

Segm. En phrases

$$\begin{bmatrix} \bar{x}_{1,1,\bullet} & \dots & \bar{x}_{1,j,\bullet} & \dots & \bar{x}_{1,N,\bullet} \\ \vdots & \ddots & \vdots & \dots & \vdots \\ \bar{x}_{i,1,\bullet} & \dots & \bar{x}_{i,j,\bullet} & \dots & \bar{x}_{i,N,\bullet} \\ \vdots & \dots & \vdots & \ddots & \vdots \\ \bar{x}_{K,1,\bullet} & \dots & \bar{x}_{K,j,\bullet} & \dots & \bar{x}_{K,N,\bullet} \end{bmatrix}$$

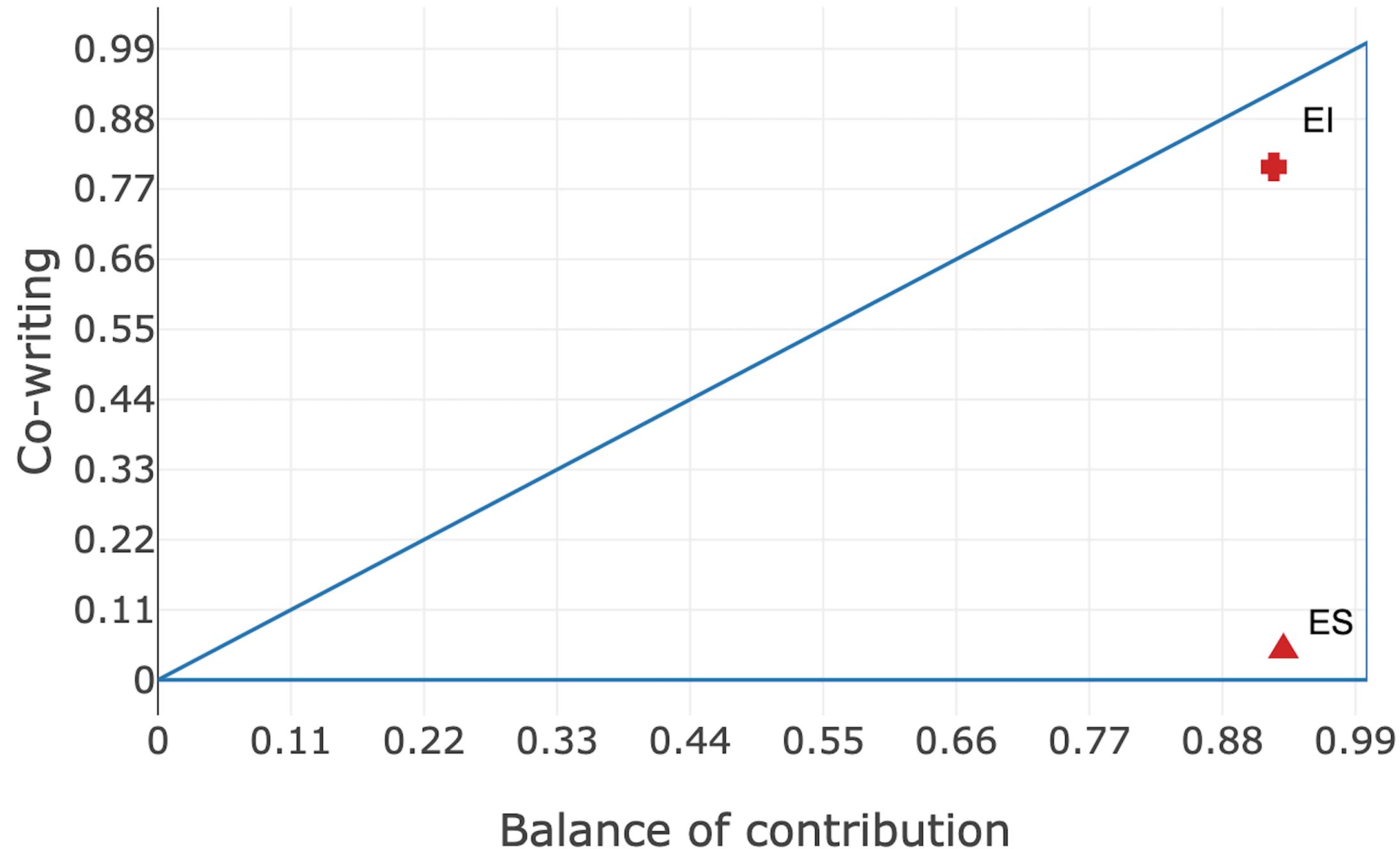
Cont. sur les phrases

$$v_j = 1 - \frac{K}{K-1} \sum_{i=1}^K \left(\bar{x}_{i,j,\bullet} - \frac{1}{K} \right)^2, \quad \forall j \in [1, N].$$

Agréger cont. Phrases

$$c(X) = \sum_{j=1}^N p_j v_j \quad \text{where} \quad p_j = \frac{n_j}{N}.$$

Stratégies de collaboration



II. Indicateurs d'écriture collaborative « non-sémantique » : bref aperçu

Limite principale

- ▶ Publication : A. M. Haddouche, Fahima Djelil, Christian Hoffmann, Nadine Mandran and Cédric d'Ham, Proposal of Indicators for Measuring Collaborative Writing in a Digital Learning Environment, Proceedings of the 15th International Conference on Computer Supported Education - Volume 2 : CSEDU, pp 495-502, 2023, Prague, Czech Republic,
- ▶ Ces deux indicateurs ne prennent pas en compte l'évolution sémantique du texte

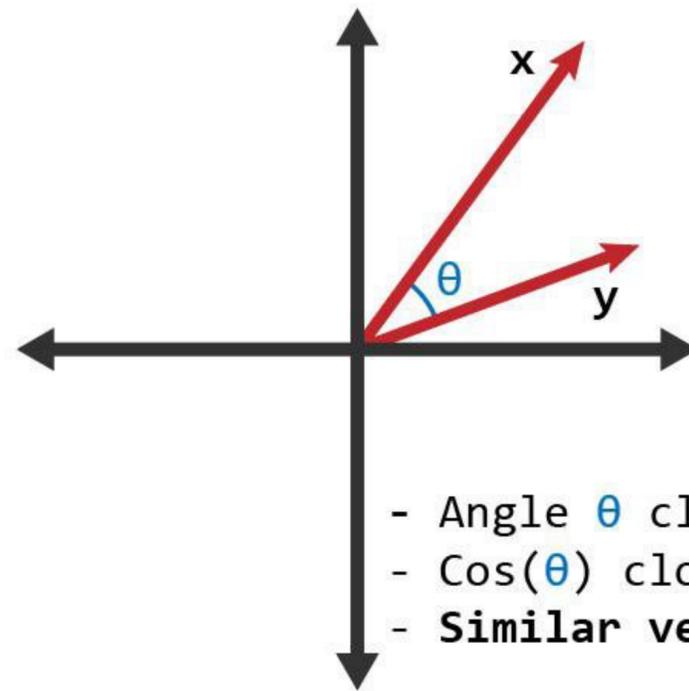
III. Indicateur d'écriture collaborative « sémantique »

Comment mesurer la différence **sémantique** entre deux documents ?

- ▶ L'idée générale est de représenter chaque document dans un vecteur numérique compréhensible par la machine
- ▶ Ensuite, pour comparer la similarité textuelle de ces deux documents, il suffit de comparer leurs deux vecteurs respectifs avec la **distance cosinus**

III. Indicateur d'écriture collaborative « sémantique »

Similarité Cosinus



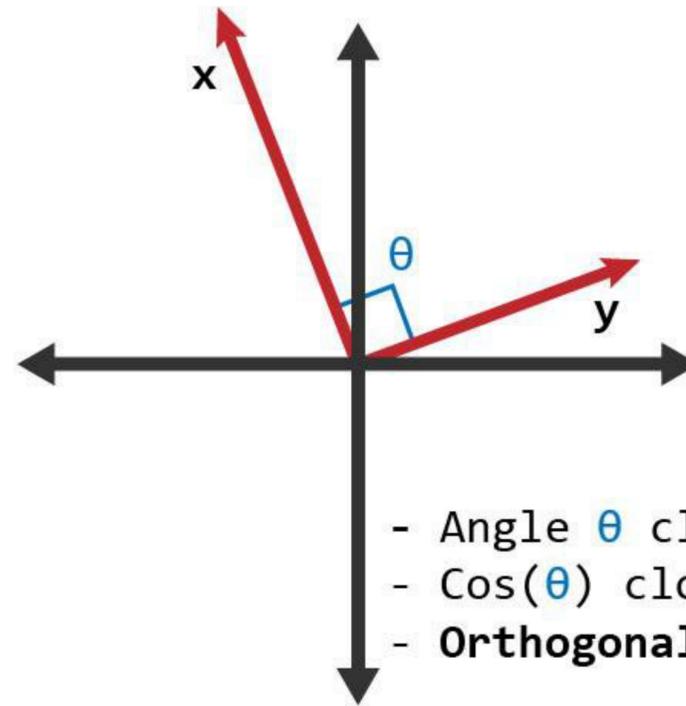
- Angle θ close to 0
- $\text{Cos}(\theta)$ close to 1
- **Similar vectors**

Similarité

Phrase x : "Les oiseaux chantent joyeusement dans le ciel bleu."

Phrase y : "Les oiseaux gazouillent dans le ciel ensoleillé. »

Distance cosinus : 0.95



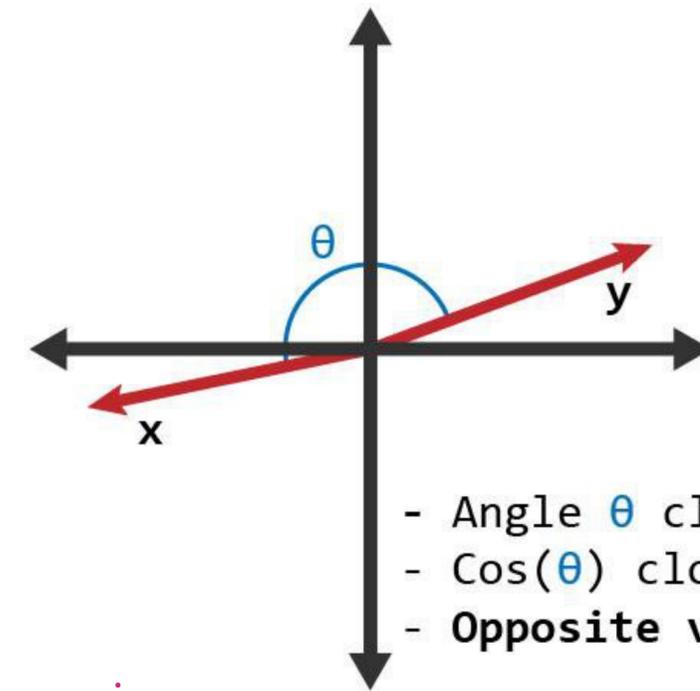
- Angle θ close to 90
- $\text{Cos}(\theta)$ close to 0
- **Orthogonal vectors**

Pas de similarité

Phrase x : "Le chat noir se repose sur le tapis."

Phrase y : "Le chien brun joue dans le jardin."

Distance cosinus : 0.12



- Angle θ close to 180
- $\text{Cos}(\theta)$ close to -1
- **Opposite vectors**

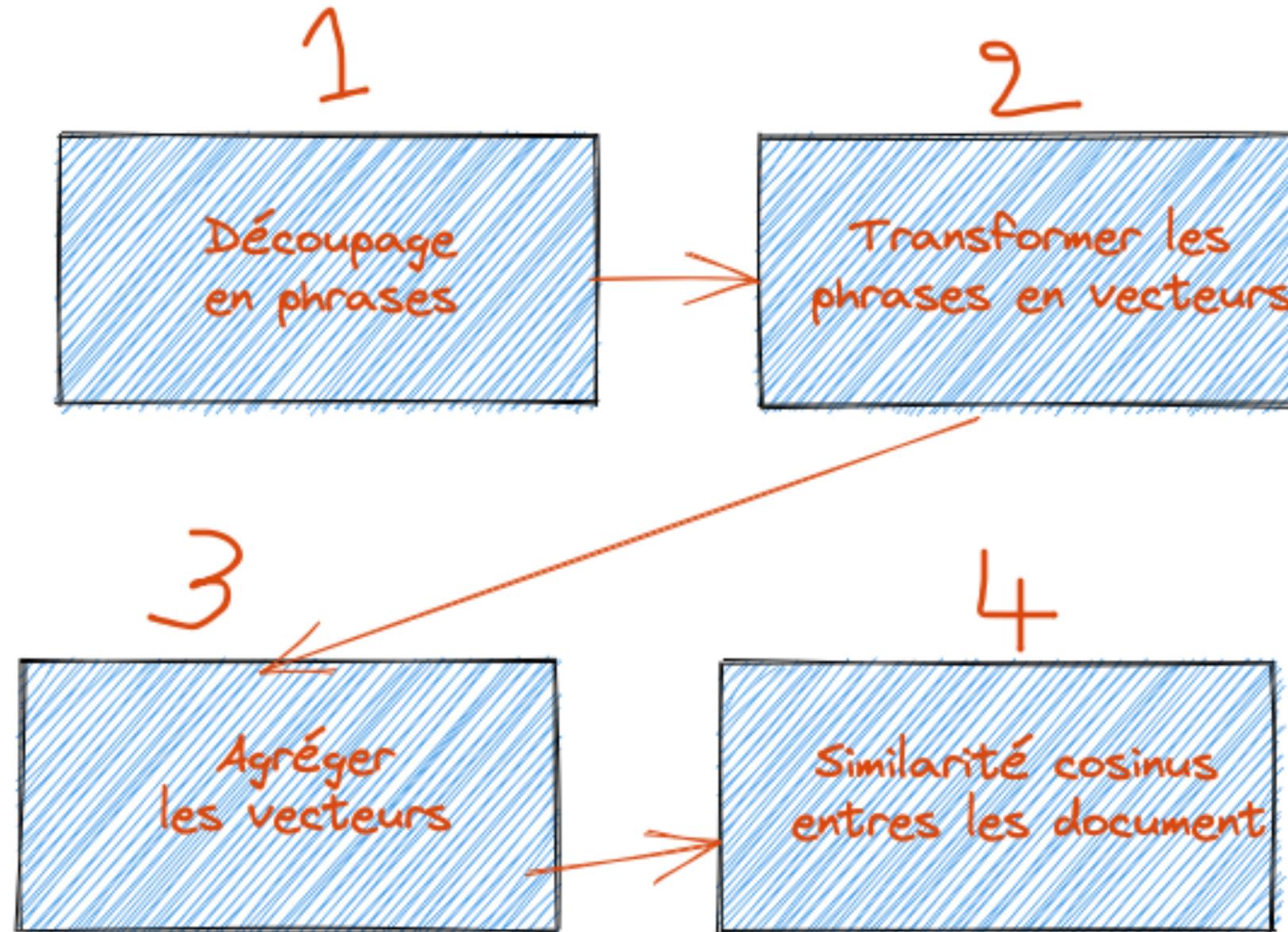
Dissimilarité

Phrase x : "La voiture rouge accélère sur l'autoroute."

Phrase y : "La voiture bleue ralentit près du feu de signalisation. »

Distance cosinus : -0.25

Etapes de construction

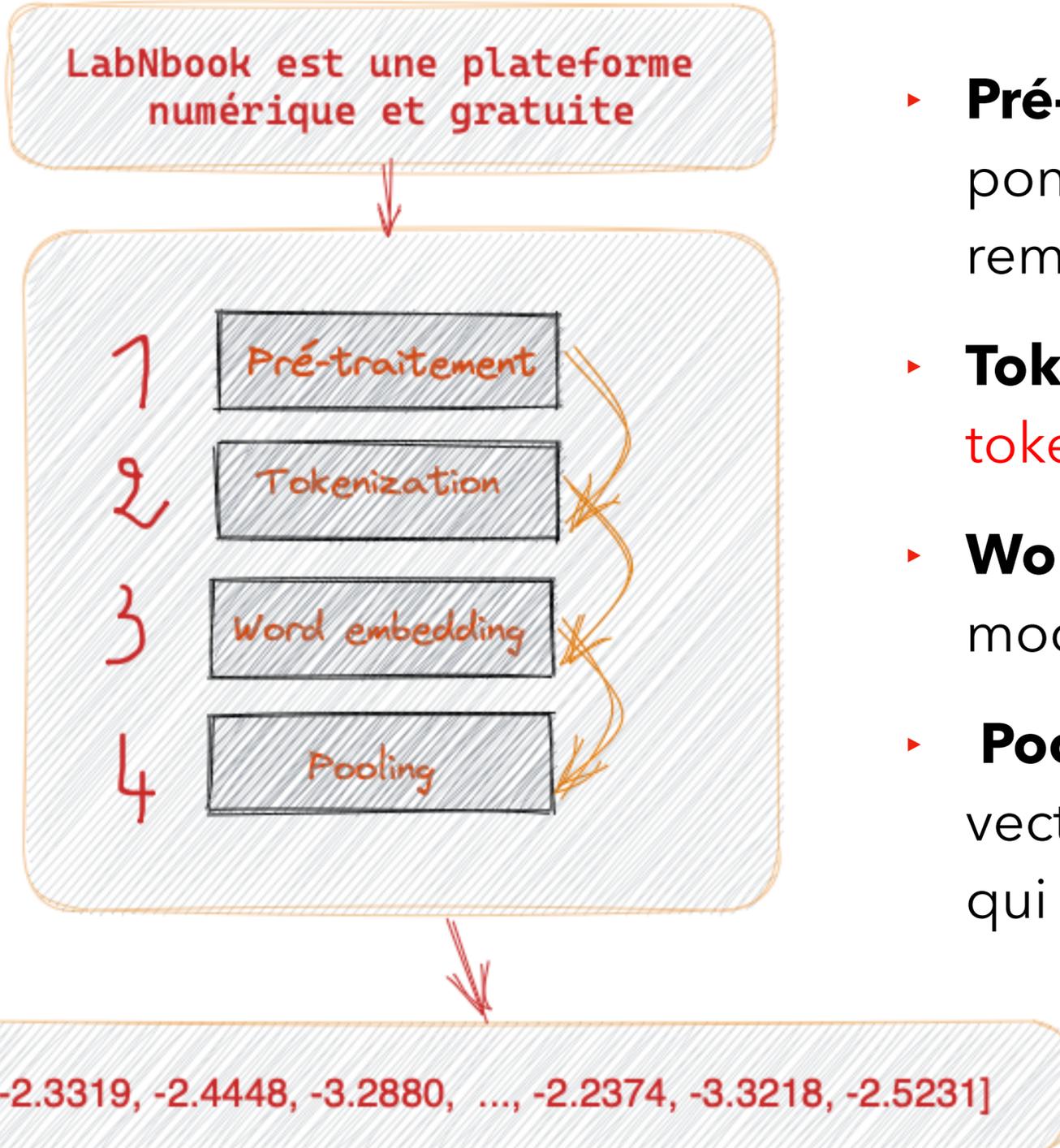


Embedding

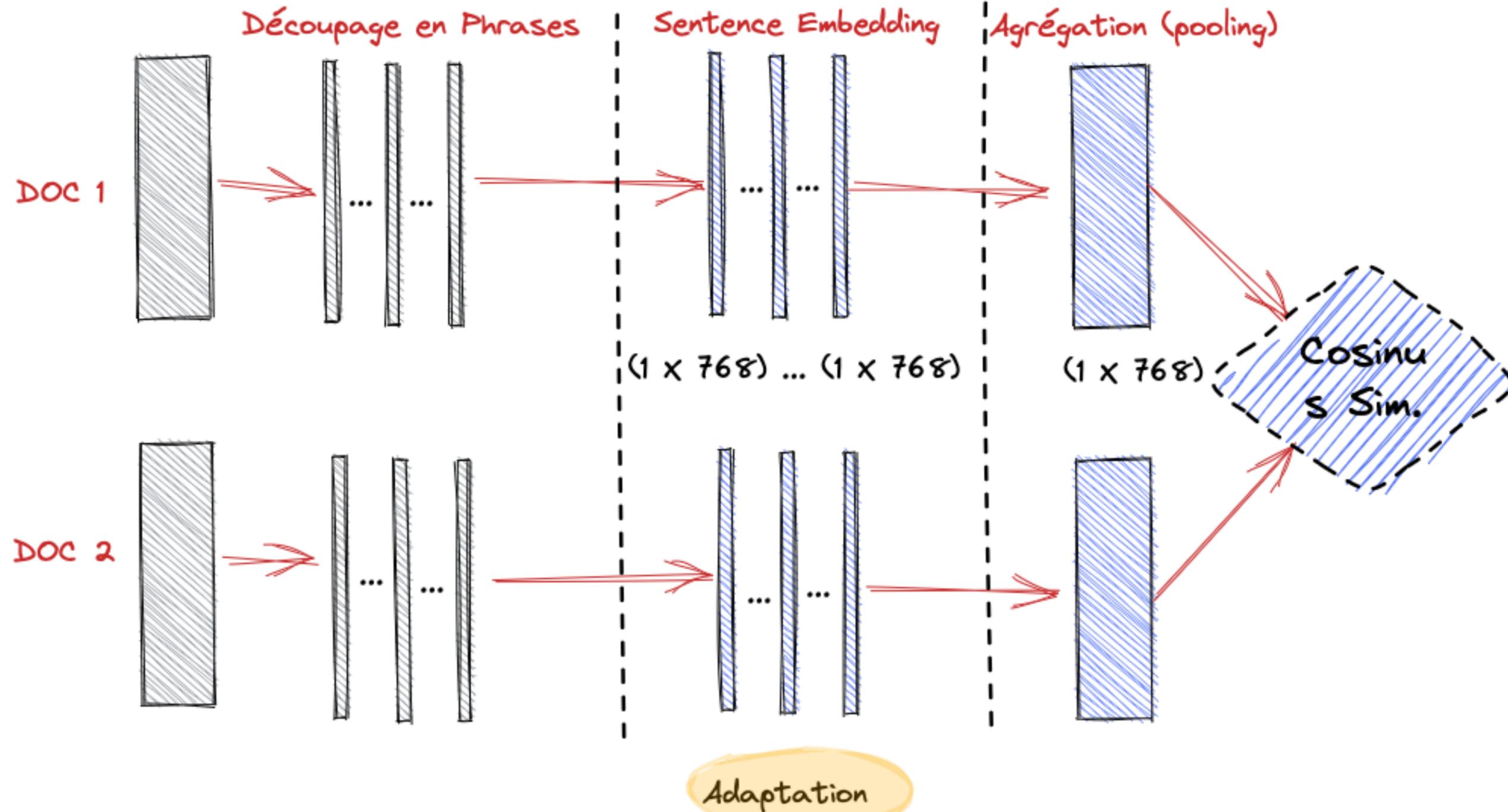
- ▶ L'application du embedding à une phrase donne une représentation vectorielle sémantique de celle-ci
- ▶ Les modèles de type **Transformers** ont une architecture qui peut être utilisée pour faire le **sentence embedding**

Embedding

- ▶ **Pré-traitement** : Mettre en minuscules, supprimer la ponctuation et les caractères spéciaux, ainsi que les mots de remplissage tels que "le", "la", "de", etc.
- ▶ **Tokenization** : Convertir nos phrases en une séquence de **tokens** en utilisant un tokenizer spécifique.
- ▶ **Word Embedding** : Mettre notre séquence de tokens dans un modèle de type Transformer pré-entraîné, tel que **BERT**.
- ▶ **Pooling** : Consiste à prendre la moyenne ou la somme des vecteurs de sortie. Cette opération nous donne un seul vecteur qui représente la phrase entière.



Processus complet



Comment construire l'indicateur ?

- ▶ A chaque changement de rédacteur on calcule la similarité cosinus
- ▶ On la transforme pour avoir des valeurs dans $[0,1]$
- ▶ On fait une moyenne cumulée à chaque version

Evolution du document

Auteur 1 : V 1



Auteur 2 : V 2



Auteur 1 : V 3



Auteur 2 : V 4

Evolution sémantique



0



1



0.5



0

Agrégation



0



1



0.75



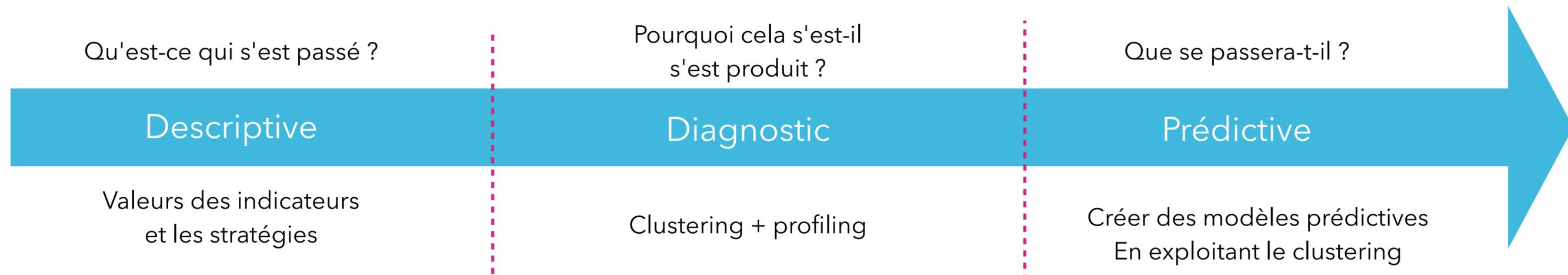
0.5

IV. Perspectives

- ▶ Comment évaluer cet indicateur ?
- ▶ Comment le combiner avec les autres indicateurs non-semantic ?

IV. Perspectives

- ▶ Quelle type d'analyse remonter à l'enseignant (Quel niveau d'indicateur pour quel usage)?



Merci pour votre attention