

Évaluation de la Robustesse des Pipelines de Recherche avec des Générateurs de Variations de Requêtes

Gustavo Penha

Arthur Câmara

Claudia Hauff

Anyes TAFOUGHALT
Racha Nadine DJEGHALI

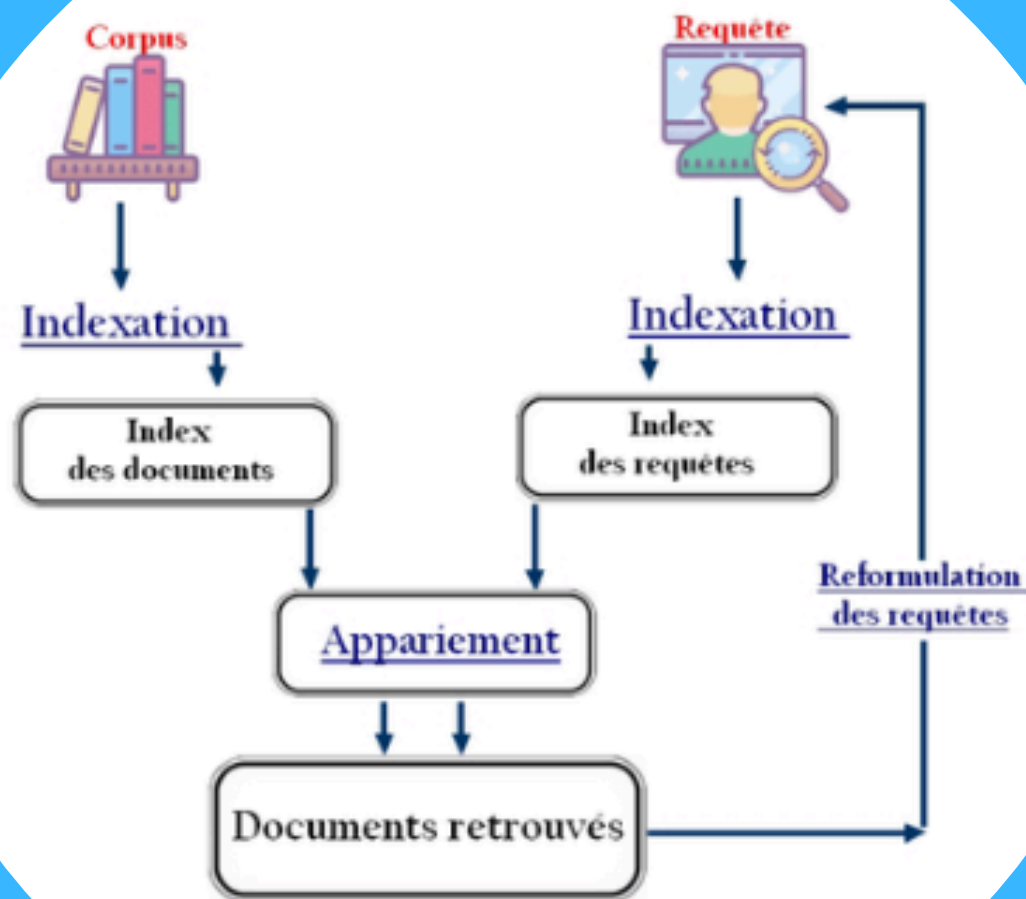
- Encadré par : Laure SOULIER



Plan

- ➊ **Méthodologie et Expérimentations**
- ➋ **Transformations de Requêtes**
- ➌ **Annotation des Requêtes Générées**
- ➍ **Évaluation et Analyse des résultats**
- ➎ **Robustesse par Catégorie de Variation/Modèle**

Introduction



Contexte

- Les transformers pré-entraînés comme BERT sont très efficaces pour les tâches de recherche d'information (IR).
- Les benchmarks IR évaluent les modèles de classement avec une seule requête pour chaque besoin d'information.

Problème



- Les utilisateurs génèrent des **variations** de requêtes **très variables** pour un même besoin d'information.

Problématique

**Comment les pipelines de recherche réagissent-elles
aux variations de requêtes qui ne changent pas la
sémantique des requêtes ?**



Méthodologie et Expérimentations

Génération automatique des variations de requêtes.

Datasets utilisés : TREC-DL-2019 et ANTIQUE.

Annotation des variations de requêtes obtenues.

Sélection des requêtes valides seulement.

Évaluation de la robustesse de la tâches RI avec différents modèles de classement.

- Modèles lexicaux traditionnels (BM25, RM3).
- Modèles de reranking neuronaux (KNRM, CKNRM).
- Modèles de reranking basées sur les transformers (EPIC, BERT, T5).

Transformations de Requêtes

Misspelling	Naturality	Ordering	Paraphrasing
<ul style="list-style-type: none">• NeighbCharSwap : Échange de caractères voisins.• RandomCharSub : Remplacement d'un caractère par un autre aléatoire.• QWERTYCharSub : Remplacement d'un caractère par un voisin du clavier QWERTY.	<ul style="list-style-type: none">• RemoveStopWords : Suppression des stopwords.• T5DescToTitle : Génération de titres de requêtes à partir de descriptions avec T5.	<ul style="list-style-type: none">• RandomOrderSwap : Échange aléatoire de deux mots.	<ul style="list-style-type: none">• BackTranslation : Traduction aller-retour via une langue pivot.• T5QQP : Génération de paraphrases de questions avec T5.• WordEmbedSynSwap : Remplacement par un synonyme selon l'embedding.• WordNetSynSwap : Remplacement par un synonyme de WordNet.

Annotation des Requêtes Générées



Validation automatique

des variations de **misspelling** et d'**ordering**, en raison de leur nature de transformations basées sur des règles.



Invalidation automatique

de toutes les transformations générant une variation **identique** à la requête d'entrée.



Annotation Manuelle

Les annotateurs annotent indépendamment les paires restantes.

Évaluation et Analyse des résultats



Efficacité (nDCG@10) des différentes méthodes face aux variations de requêtes

- ANTIQUE

	name	BM25	RM3	KNRM	CKNRM	EPIC	BERT	T5
0	original_query	0.228598	0.217015	0.195348	0.206472	0.265964	0.298741	0.333343
1	NeighbCharSwap	0.155925	0.146874	0.149475	0.144357	0.184004	0.219756	0.250900
2	QWERTYCharSub	0.161297	0.152542	0.153281	0.155270	0.192009	0.228915	0.265228
3	RandomCharSub	0.162307	0.159323	0.150131	0.147622	0.187947	0.218040	0.245766
4	RemoveStopWords	0.226959	0.216080	0.225525	0.215285	0.269284	0.278089	0.319982
5	T5DescToTitle	0.167330	0.164611	0.172450	0.167238	0.200639	0.213728	0.239296
6	RandomOrderSwap	0.228587	0.216919	0.195337	0.197762	0.266102	0.296807	0.325507
7	BackTranslation	0.161758	0.154594	0.135938	0.143829	0.203180	0.232330	0.258100
8	T5QQP	0.220146	0.206255	0.188036	0.195704	0.261390	0.287625	0.321404
9	WordEmbedSynSwap	0.175915	0.171544	0.172996	0.168927	0.213856	0.246034	0.281352
10	WordNetSynSwap	0.179095	0.175000	0.175970	0.176346	0.211673	0.245963	0.273431

Misspelling

Naturality

Ordering

Paraphrasing

- TREC_DL_2019

	name	BM25	RM3	KNRM	CKNRM	EPIC	BERT	T5
0	original_query	0.479540	0.515595	0.496959	0.493117	0.624047	0.628664	0.699780
1	NeighbCharSwap	0.274701	0.274816	0.317241	0.307987	0.389341	0.422847	0.494417
2	QWERTYCharSub	0.243548	0.250378	0.275153	0.296539	0.349635	0.388942	0.446060
3	RandomCharSub	0.231450	0.234741	0.222502	0.226251	0.294983	0.316191	0.396311
4	RemoveStopWords	0.477754	0.511339	0.489481	0.475599	0.621385	0.625341	0.686159
5	T5DescToTitle	0.421481	0.434432	0.394968	0.392752	0.506071	0.537398	0.571683
6	RandomOrderSwap	0.479540	0.515595	0.496959	0.470756	0.622721	0.612984	0.697020
7	BackTranslation	0.396437	0.419520	0.376331	0.360486	0.530082	0.559598	0.605816
8	T5QQP	0.472212	0.504274	0.437429	0.460877	0.604045	0.629245	0.704520
9	WordNetSynSwap	0.348766	0.365042	0.359142	0.360497	0.449044	0.497299	0.545733
10	WordEmbedSynSwap	0.353017	0.353864	0.380731	0.368002	0.474884	0.501280	0.560259

- Une baisse significative de l'efficacité est observée sur TREC-DL-2019 et ANTIQUE.
- Aucune variation de requête n'améliore l'efficacité par rapport à la requête originale.
- Les pipelines de RI ne sont pas robustes aux variations de requêtes.

Explication des Baisses d'Efficacité des Ordonnanceurs Neuronaux

Défaillance de l'ordonnanceur lors de la première phase

Étant incapable de récupérer suffisamment de documents pertinents pour que les ordonnanceurs neuronaux les reclassent.

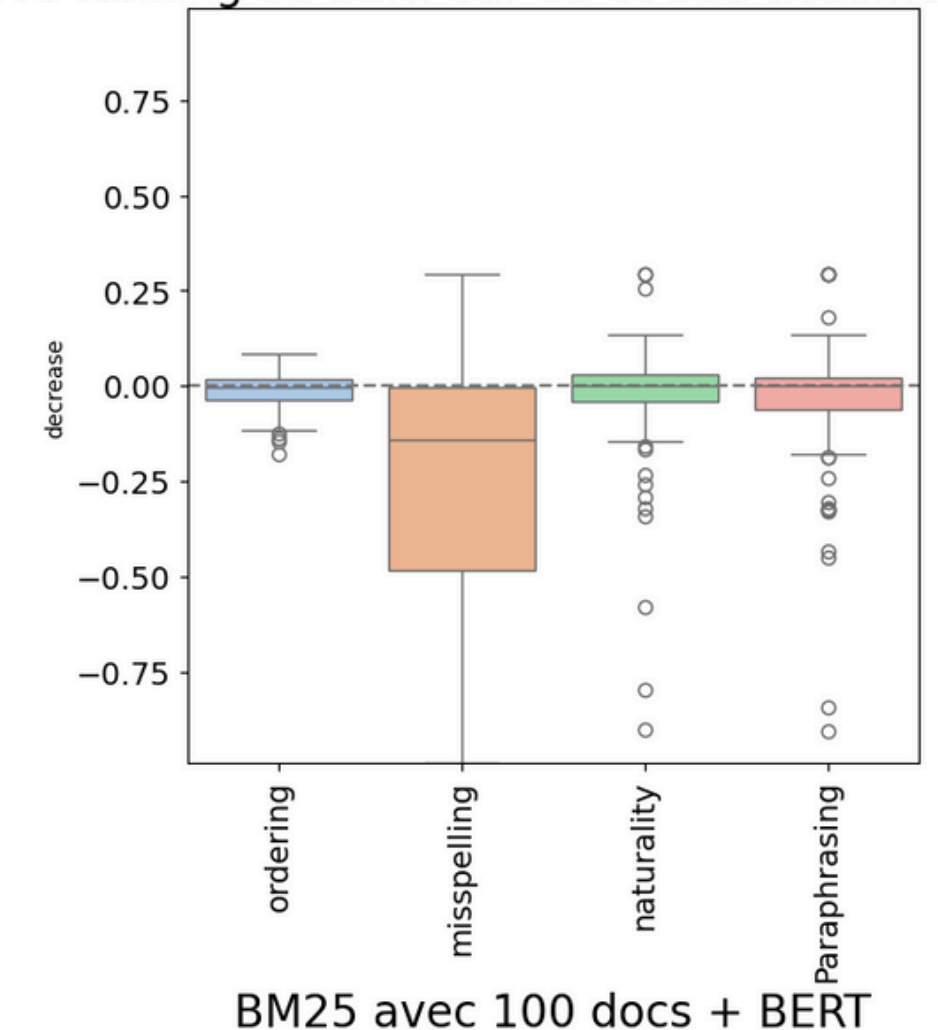
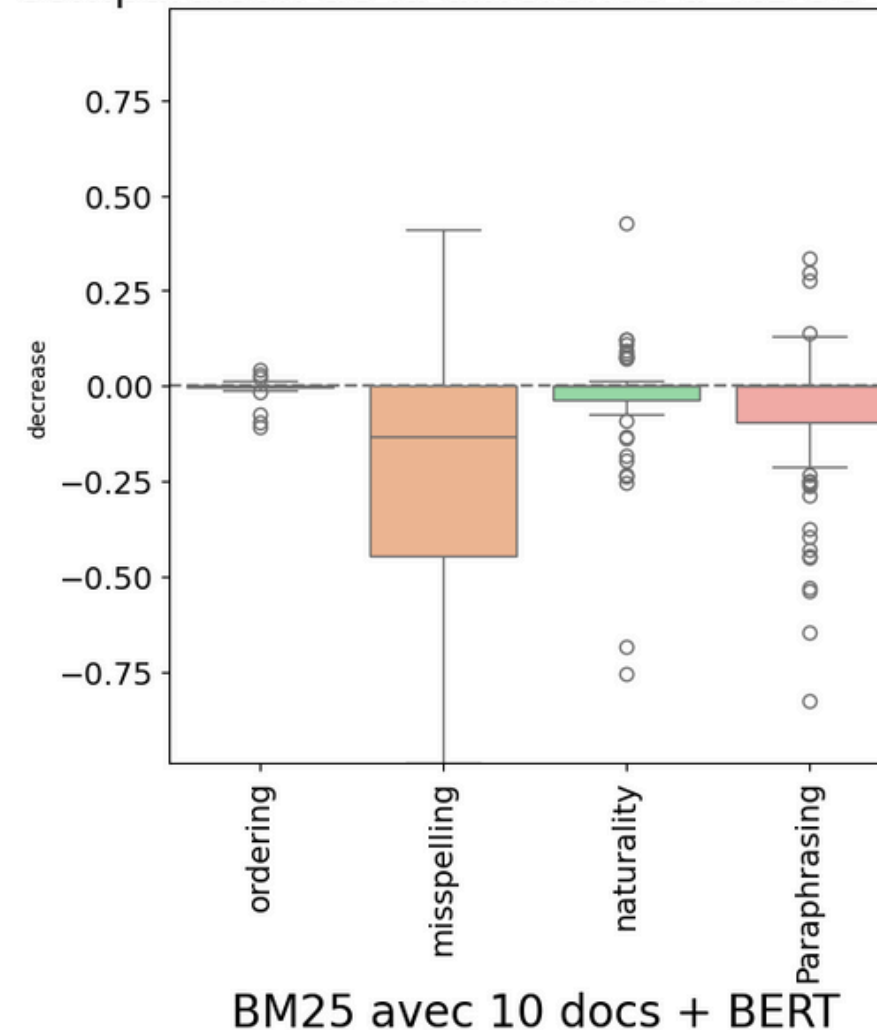
Impacts des documents non jugés

Variations de requêtes entraînant des documents non jugés hautement classés par les pipelines de récupération, jugés non pertinents dans l'évaluation standard.

Défaillance de l'ordonnanceur lors de la première phase

- Augmentation du seuil de reranking n'améliore pas significativement la performance (ex. BM25 avec Recalls @10, @100).
- Même avec un plus grand nombre de documents pertinents pour le reranker, la performance du ranker de première étape reste insuffisante pour gérer les variations de requêtes.

Comparaison de la différence d' NDCG Cut 10 entre un re-ranking de BERT sur 10 et 100 documents



Impact des documents non jugés

- Réaliser la récupération de première étape sur la requête originale et à utiliser les variations de requêtes uniquement pour les modèles de reranking.
- Les résultats montrent que des baisses d'efficacité persistent, bien que de moindre ampleur.

1) Avec requêtes originales en first stage

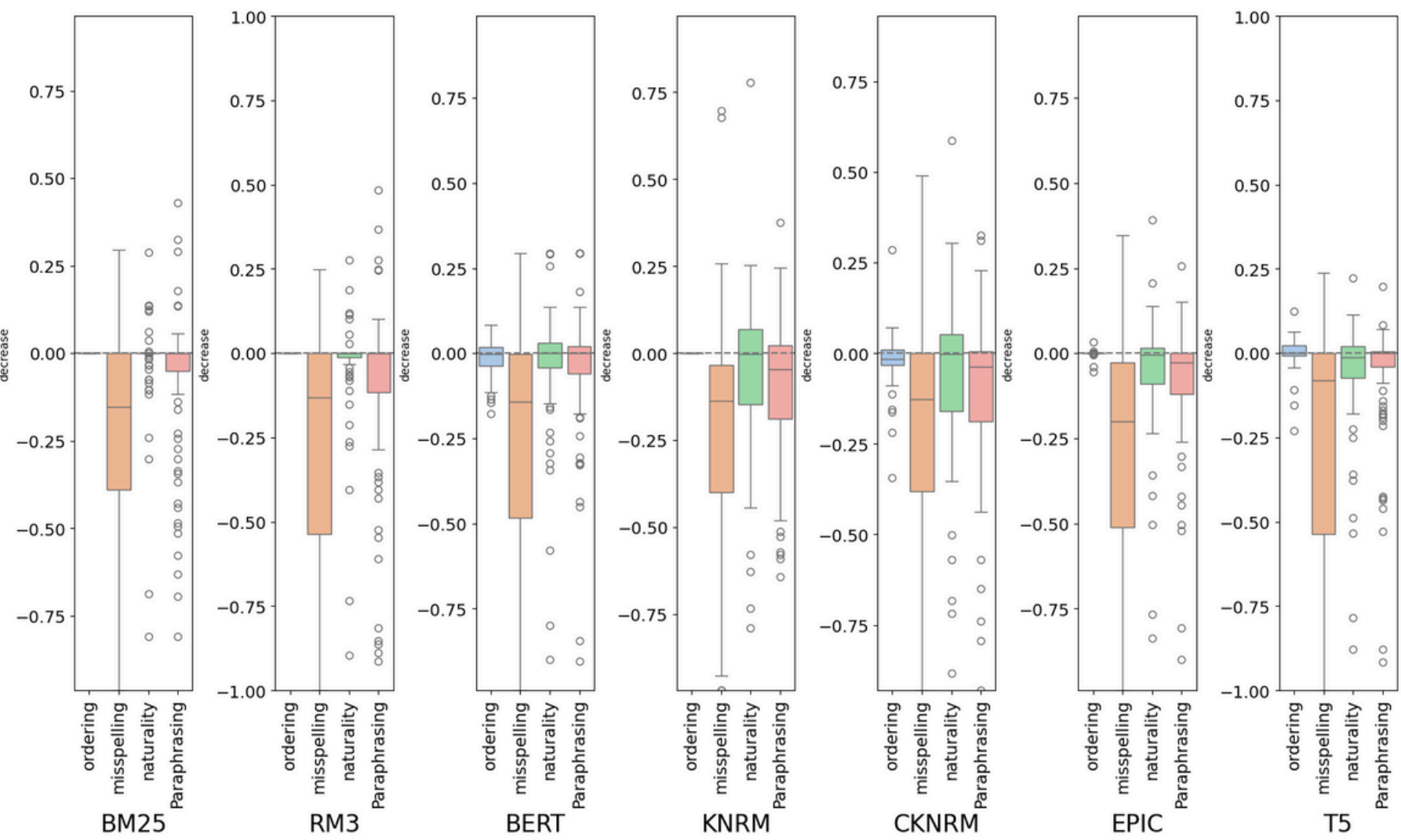
	name	ndcg_cut_10	EPIC
0	BM25+onoriginal+EPIC	0.624047	0.624047
1	BM25+onoriginal+EPIC+NeighbCharSwap	0.536365	0.389341
2	BM25+onoriginal+EPIC+QWERTYCharSub	0.536365	0.349635
3	BM25+onoriginal+EPIC+RandomCharSub	0.536365	0.294983
4	BM25+onoriginal+EPIC+RemoveStopWords	0.536365	0.621385
5	BM25+onoriginal+EPIC+T5DescToTitle	0.536365	0.506071
6	BM25+onoriginal+EPIC+RandomOrderSwap	0.536365	0.622721
7	BM25+onoriginal+EPIC+BackTranslation	0.536365	0.530082
8	BM25+onoriginal+EPIC+T5QQP-base	0.536365	0.604045
9	BM25+onoriginal+EPIC+WordNetSynSwap	0.536365	0.449044
10	BM25+onoriginal+EPIC+WordEmbedSynSwap	0.536365	0.474884

2) Qu'avec les variations

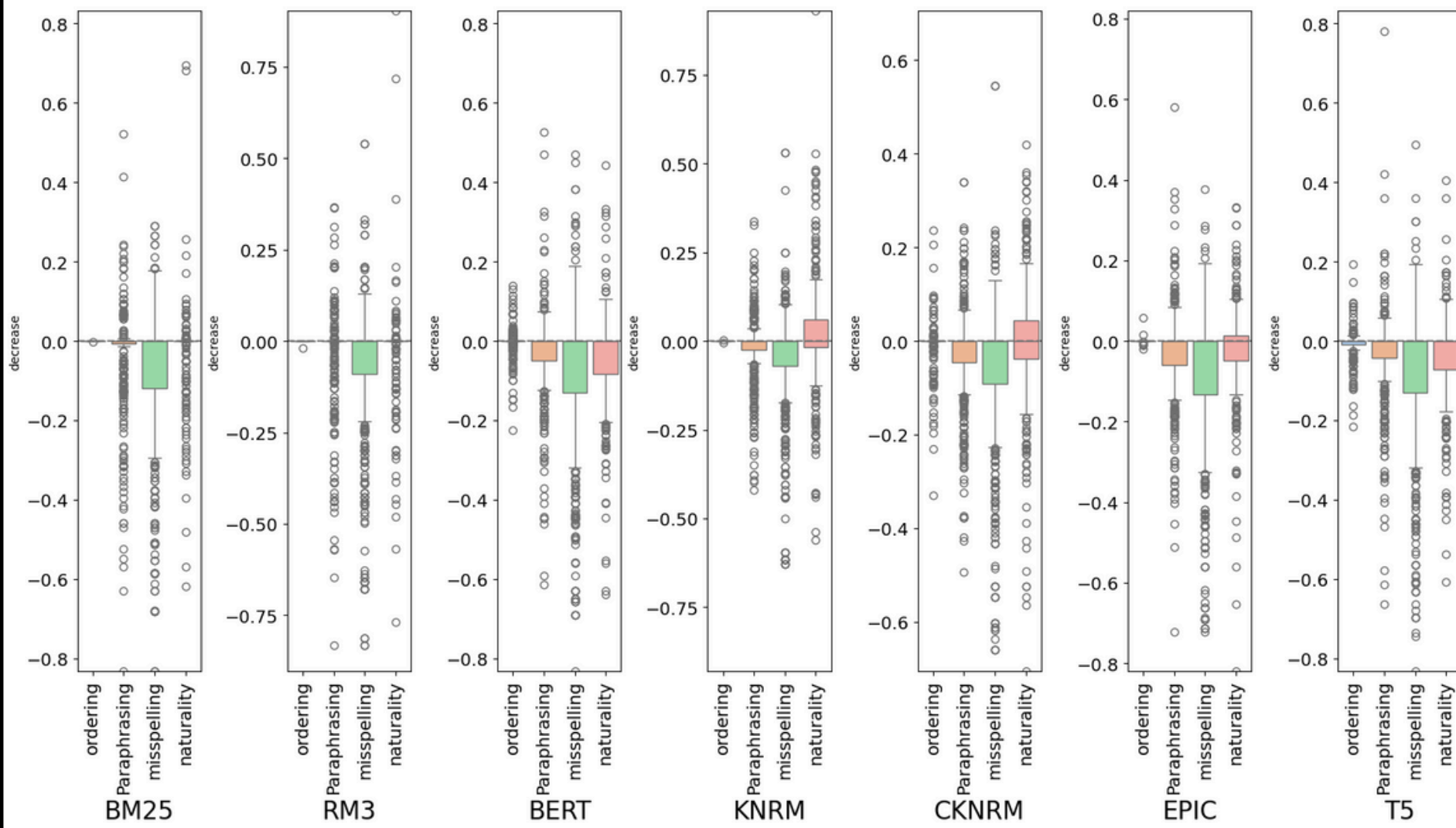
Robustesse par Catégorie de Variation

Distribution de la variation de nDCG@10 lors du remplacement de la requête originale par les méthodes de chaque catégorie

- Sur les données TREC

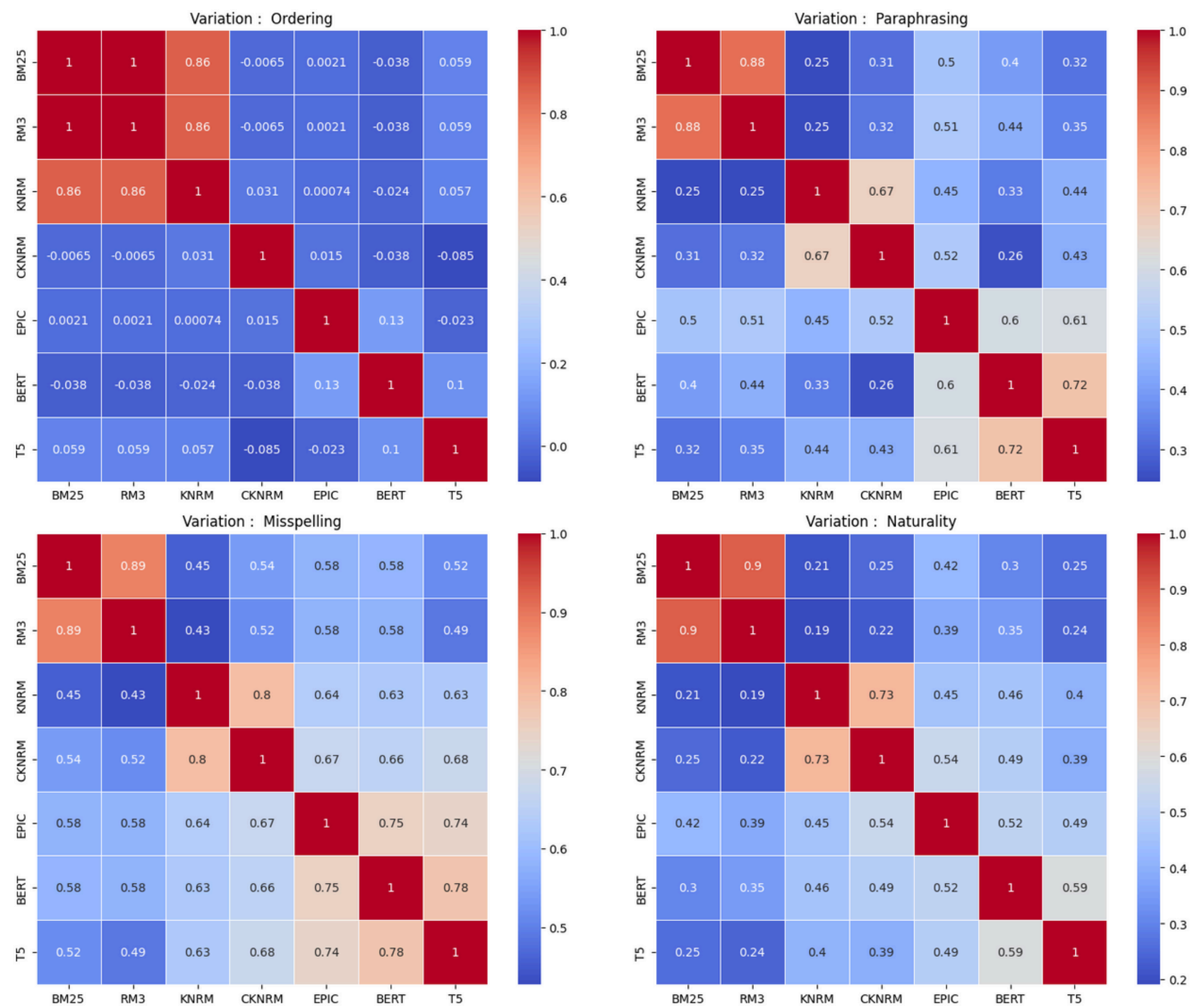


- Sur les données ANTIQUE



Robustesse par Catégorie de Modèle

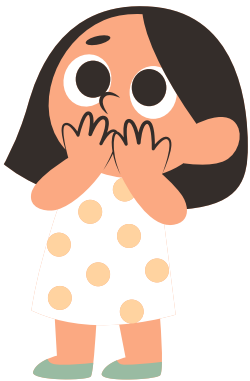
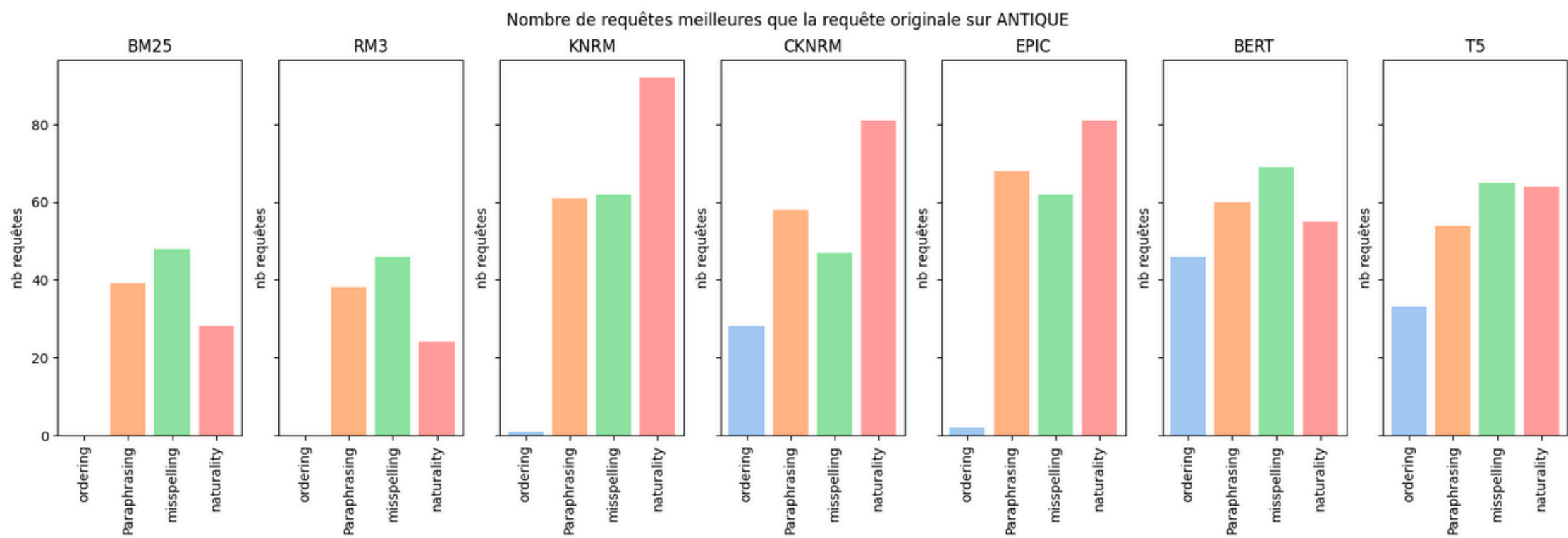
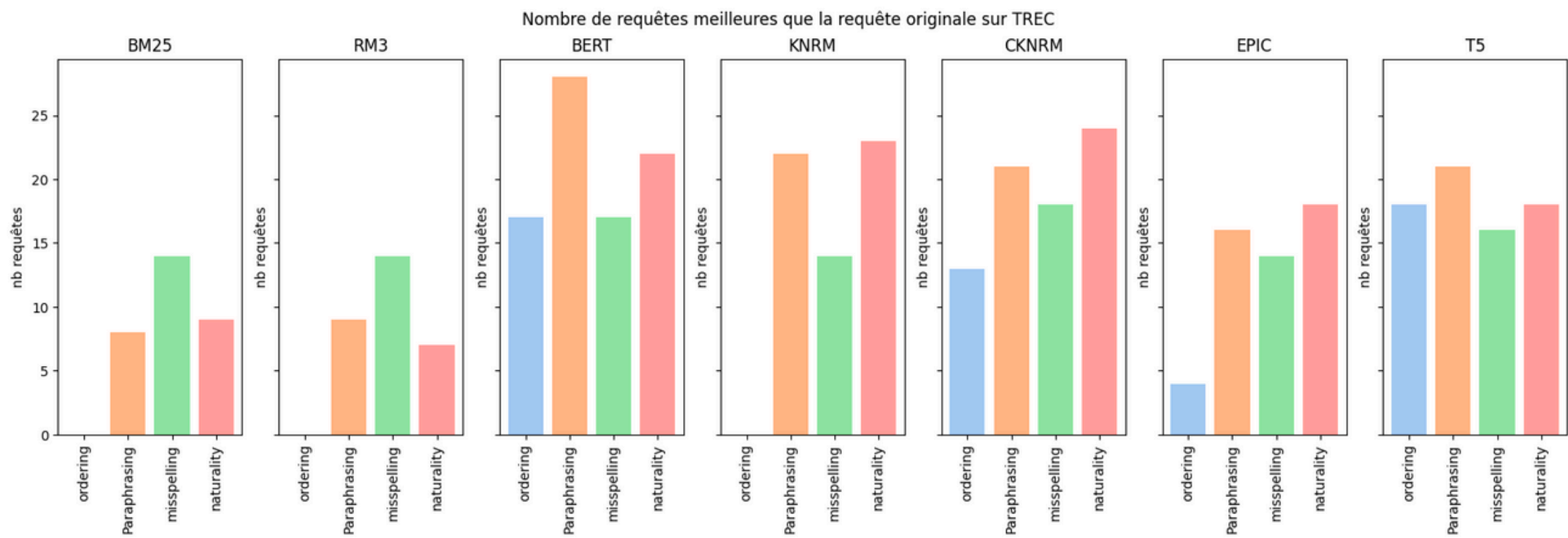
Corrélation de Pearson entre la variation de nDCG@10 des modèles de classement



The slide features two large, solid blue circles. One is in the bottom-left corner, and the other is in the top-right corner, both partially cut off by the edges of the frame.

Meilleur pour la fin ?

Quelques Améliorations !!



Conclusion

- Les pipelines de récupération montrent une baisse significative de performance avec des variations de requêtes.
- Les collections IR doivent inclure des variations de requêtes pour mieux évaluer la robustesse des modèles.

Perspectives

- Intégration plus poussée des variations dans l'entraînement des modèles de RI.
- Exploration de techniques avancées telles que la fusion des variations de requêtes pour maximiser les avantages tout en minimisant les inconvénients.

The slide features two large, solid blue circles. One circle is positioned in the bottom-left corner, and the other is in the top-right corner, both partially cut off by the edges of the frame.

Merci de Votre Attention !