

# SPQR@Deft2023: Résolution automatique de QCM médicaux à partir de corpus de domaine et de mesures de similarité

---

**Auteurs** : Julien Bezançon, Toufik Boubehziz, Corina Chutaux, Oumaima Zine, Laurie Acensio, Caroline Koudoro-Parfait, Andrea Briglia, Gaël Lejeune

**Présentation** : Julien Bezançon, Corina Chutaux

05/06/2023

STIH (Sens Texte Informatique Histoire), OB TIC (Observatoire des Textes, des Idées et des Corpus), CERES (Centre d'expérimentation en méthodes numériques pour les recherches en Sciences Humaines et Sociales), Sorbonne Université

Introduction

Classification des questions

Construction des corpus de référence

Méthode de détection des termes médicaux

Expériences et résultats

Conclusion

# Introduction

---

Méthode de similarité :

- répondre à la tâche d'identification des réponses correctes parmi les cinq réponses possibles du QCM de pharmacologie
- identifier le nombre de réponses correctes pour une question nous avons testé une méthode de similarité

Postulat de base = chercher comment nous pouvons montrer que les bonnes réponses à une question étaient similaires à des phrases trouvées dans un texte de référence

# Classification des questions

---

Trois types des classifications effectuées :

1. La première visait à décomposer les questions en trois parties d'un point de vue de leur forme : [P1][P2][P3]
  - [P1] était l'entête
  - [P2] le type de question attendue (correcte, exacte, inexacte etc.)
  - [P3] comportait le contexte/ l'information pharmacologique
2. La deuxième analysait :
  - **Les composants morphologiques** : adverbes et verbes récurrents dans la formulation des questions : parmi, donner, cocher etc. ; négations ; ou les termes utilisés dans l'affirmation ou la négation comme (in)exacte, juste, fausse etc.
  - **Les composants sémantiques** : indication implicite ou explicite sur le nombre des réponses possibles ; informations médicales ;

# Classification des questions

3. La troisième répertoriait les questions en s'appuyant sur leur ponctuation :

Terminaison de la question	Nombre de questions
" ?" (point d'interrogation)	1147
" :" (deux points)	766
" " (pas de ponctuation finale)	118
... (ellipse)	78
" ." (point)	52
" ," (virgule)	10

**Table 1** – Nombre de questions par type de terminaison

# Construction des corpus de référence

---

# Corpus Feedback

À partir de ces classifications nous avons abouti au Corpus Feedback.

Ce corpus a été constitué en transformant les corpus d'entraînement en texte composé de propositions assertives.

Trois transformations automatiques réalisées :

1. Concaténation des questions avec réponses correctes
2. Transformation des phrases interrogatives en subordonnées relatives
3. Paraphrase

# Corpus Feedback Bis

Deux types de questions ont été traitées : celles comportant les termes « exact » et « inexactes » (soit 46% des questions du jeu d'entraînement).

Deux transformations ont découlé de cette classification :

1. Transformation des questions en assertions avec des expressions régulières pour les deux classes
2. Concaténation de chaque assertion obtenue à l'étape 1 avec les réponses correctes

Différence entre corpus Feedback et Feedback Bis - à chaque question on attribue une assertion distincte pour chaque réponse correcte, tandis que dans le corpus feedback nous avons ajouté des conjonction de coordination pour concevoir des phrases exhaustives.

# Corpus ChatGPT

Pour enrichir ces deux corpus nous avons constitué le Corpus Chat GPT.

Interrogation de Chat GPT pour élargir la couverture du corpus Feedback. Chaque assertion du corpus Feedback a été envoyé au modèle sous forme de requête CURL via L'API qui a renvoyé un ou plusieurs paragraphes complétant l'assertion initiale.

Exemple :

*"Le crack est une forme de cocaïne" : "le crack est une forme de cocaïne qui est transformée en une substance solide et cristalline, généralement fumée plutôt que sniffée. Le crack est considéré comme une drogue très addictive et dangereuse en raison de son effet rapide et intense sur le système nerveux central."*

# Corpus Merck

Basé sur le Manuel Merck disponible en ligne dont le contenu a semblé pertinent comparé à notre jeu d'entraînement. Le contenu du manuel a été récupéré et ajouté au Corpus ALL IN ONE

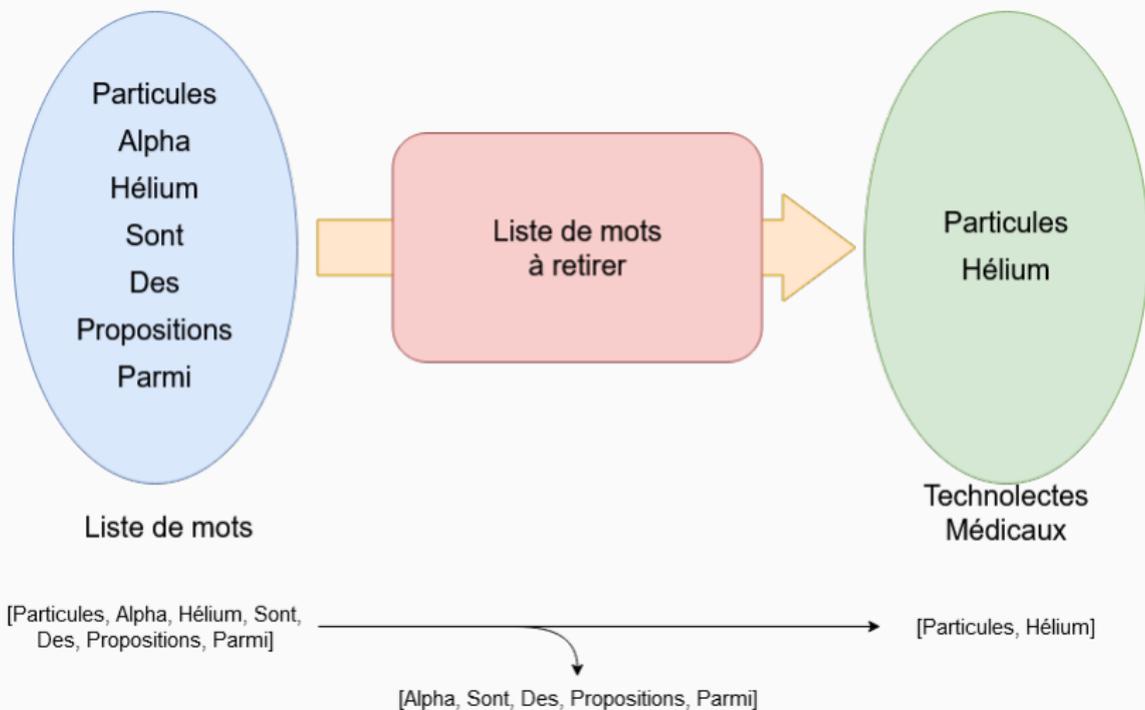
# Corpus All In One

Corpus ALL IN ONE a représenté notre objectif dès le départ, il est venu rassembler tous nos corpus en un seul. Le but de la construction de corpus a été de nourrir notre système de similarité avec des termes techniques et du contexte.

# Méthode de détection des termes médicaux

---

# Traitement des corpus



**[Q] Parmi les éléments suivants, quel est celui qui n'entre pas dans l'uréogénèse ?**

$R_1$  :  $CO_2$

$R_2$  :  $NH_3$

$R_3$  : Valine

$R_4$  : ATP

$R_5$  : Ornithine

**[Q]** Parmi les éléments suivants, quel est celui qui n'entre pas dans l'uréogénèse ?

$R_1$  :  $CO_2$

$R_2$  :  $NH_3$

$R_3$  : Valine

$R_4$  : ATP

$R_5$  : Ornithine

$[Q] + [R_1]$ ,  $[Q] + [R_2]$ ,  $[Q] + [R_3] \dots$

**[Q] Parmi les éléments suivants, quel est celui qui n'entre pas dans l'uréogénèse ?**

$R_1$  : *CO<sub>2</sub>*

$R_2$  : *NH<sub>3</sub>*

$R_3$  : *Valine*

$R_4$  : *ATP*

$R_5$  : *Ornithine*

$[Q] + [R_1]$ ,  $[Q] + [R_2]$ ,  $[Q] + [R_3] \dots$

**Récupération des technolectes médicaux sur les paires créées !**

## **Expériences et résultats**

---

## Recherche des bonnes réponses

$[Q + R_1]$
-------------

$[Q + R_2]$
-------------

$[Q + R_3]$
-------------

$[Q + R_4]$
-------------

$[Q + R_5]$
-------------

$[Q + R_1]$
-------------

$[Q + R_2]$
-------------

$[Q + R_3]$
-------------

$[Q + R_4]$
-------------

$[Q + R_5]$
-------------

$[Q + R_1]$
-------------

$[Q + R_2]$
-------------

$[Q + R_3]$
-------------

$[Q + R_4]$
-------------

$[Q + R_5]$
-------------

technolectes phrase 1
-----------------------

technolectes phrase 2
-----------------------

technolectes phrase 3
-----------------------

technolectes phrase 4
-----------------------

technolectes phrase 5
-----------------------

technolectes phrase 6
-----------------------

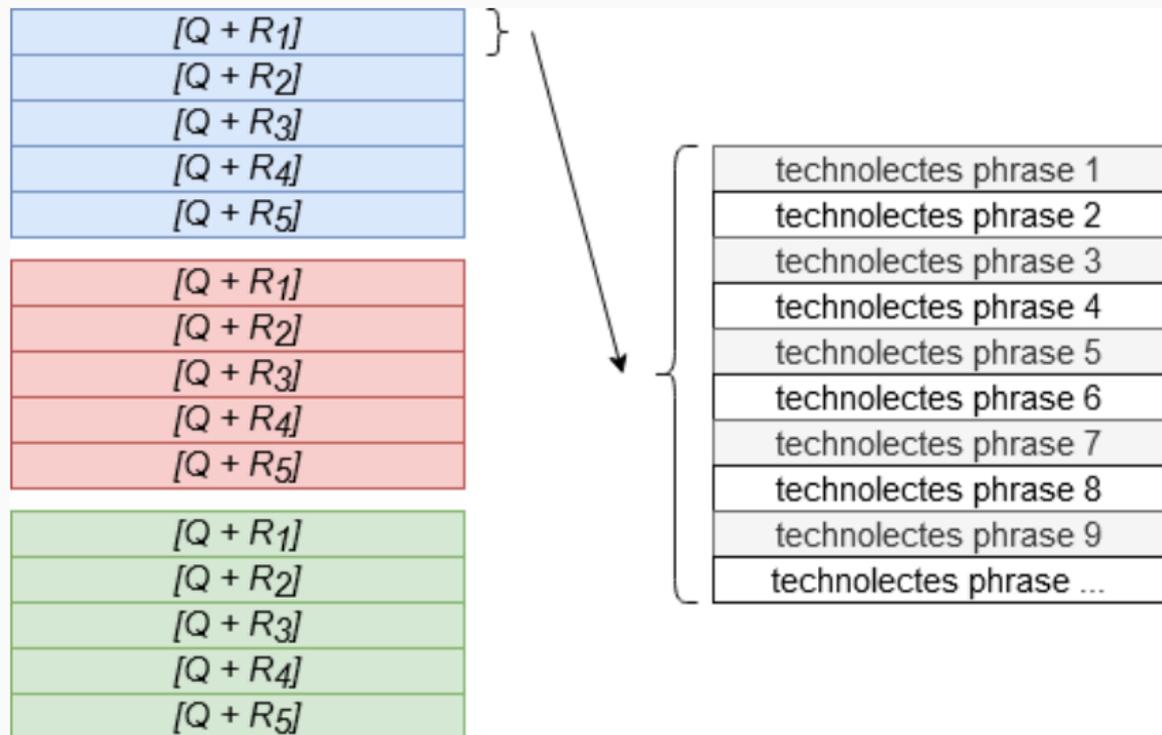
technolectes phrase 7
-----------------------

technolectes phrase 8
-----------------------

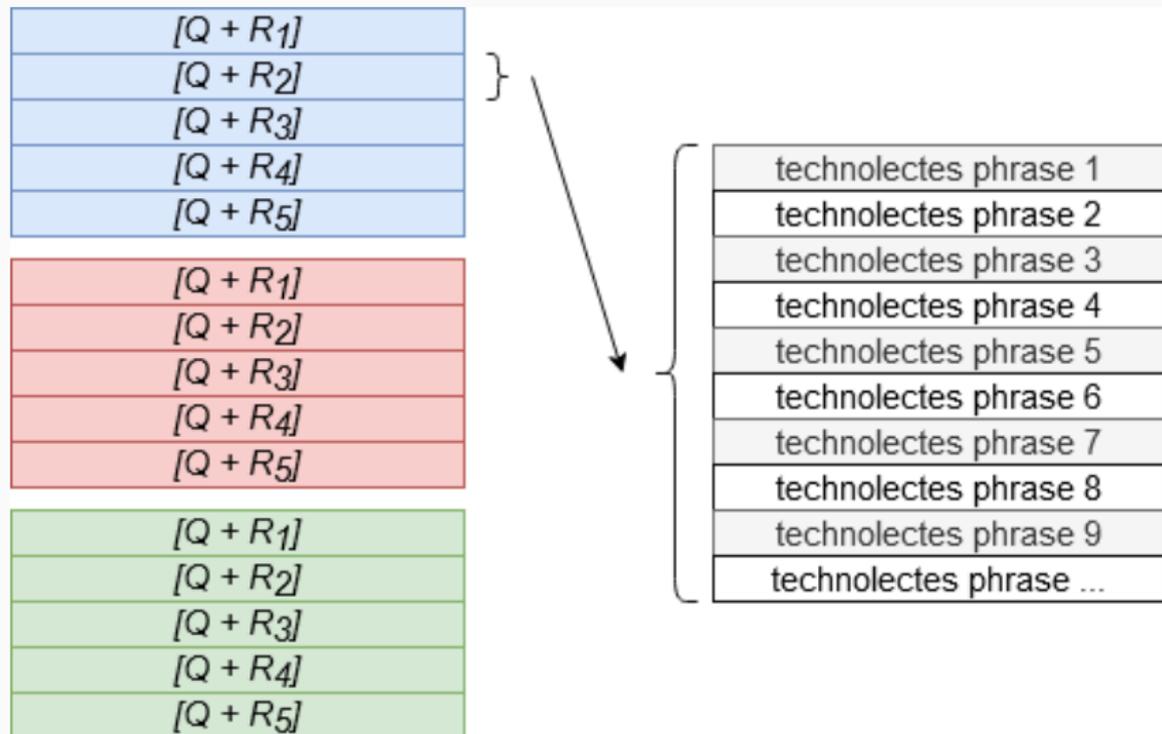
technolectes phrase 9
-----------------------

technolectes phrase ...
-------------------------

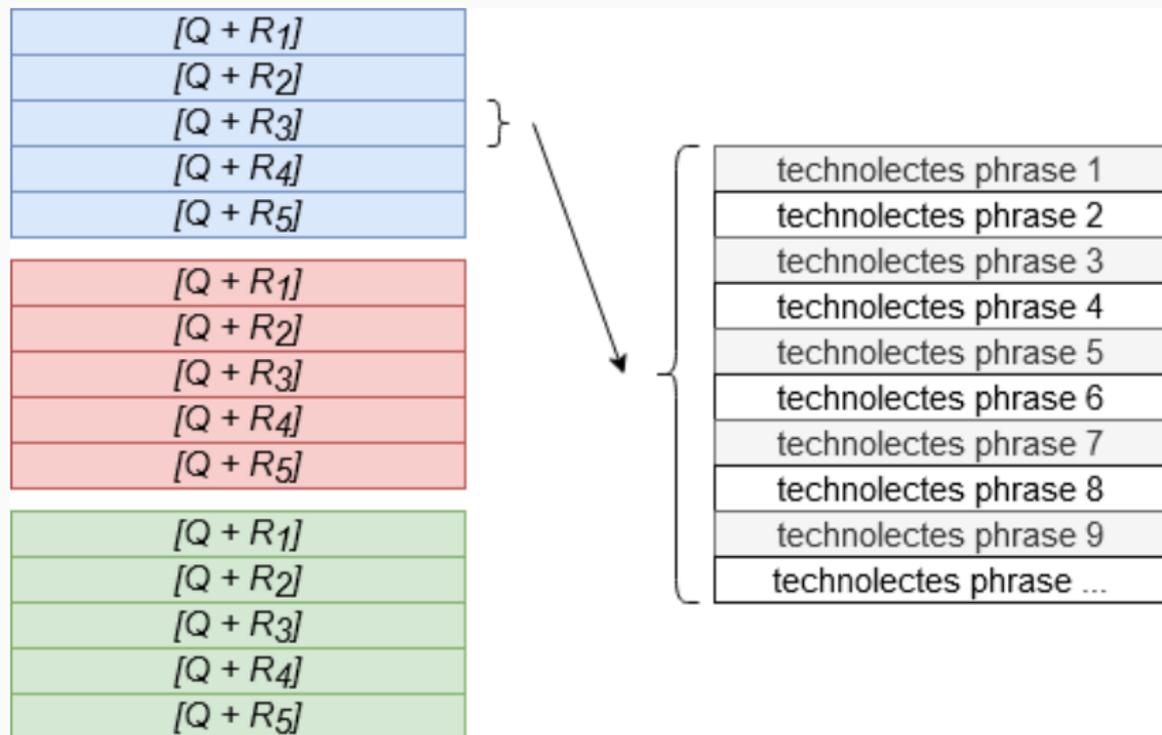
## Recherche des bonnes réponses



## Recherche des bonnes réponses



## Recherche des bonnes réponses



**Plusieurs méthode afin d'isoler de potentielles bonnes réponses :**

**Plusieurs méthode afin d'isoler de potentielles bonnes réponses :**

- Recherche par seuil (BYSEUIL)

**Plusieurs méthode afin d'isoler de potentielles bonnes réponses :**

- Recherche par seuil (BYSEUIL)
- Recherche par maximum (BYMAX)

**Plusieurs méthode afin d'isoler de potentielles bonnes réponses :**

- Recherche par seuil (BYSEUIL)
- Recherche par maximum (BYMAX)
- Fusion des deux méthodes précédentes (BYFUSION)

**Plusieurs méthode afin d'isoler de potentielles bonnes réponses :**

- Recherche par seuil (BYSEUIL)
- Recherche par maximum (BYMAX)
- Fusion des deux méthodes précédentes (BYFUSION)

**Nous ne conservons pas la méthode BYSEUIL, pouvant ne renvoyer aucun résultat !**

## Résultats sur le jeu d'entraînement

Hamming	Paramètres	EMR	Corpus
<b>0.548</b>	BYFUSION_2-3_char_cosine_0.8	0.377	All In One
0.541	BYFUSION_3-3_char_cosine_0.8	<b>0.388</b>	All In One
0.532	BYFUSION_3-3_char_cosine_0.7	0.331	All In One
0.532	BYFUSION_2-3_char_cosine_0.9	0.373	All In One
0.518	BYFUSION_2-3_char_cosine_0.7	0.252	All In One
0.505	BYFUSION_3-3_char_cosine_0.6	0.197	All In One
0.486	BYFUSION_2-3_char_cosine_0.6	0.094	All In One
0.475	BYFUSION_3-3_char_cosine_0.8	0.352	Feedback bis
0.468	BYFUSION_2-3_char_cosine_0.8	0.332	Feedback bis
0.462	BYFUSION_3-3_char_cosine_0.9	0.258	All In One

## Résultats sur le jeu de développement

Hamming	Paramètres	EMR	Corpus
<b>0.303</b>	BYFUSION_2-3_char_cosine_0.6	0.051	All In One
0.293	BYFUSION_2-3_char_cosine_0.6	0.112	Feedback bis
0.279	BYFUSION_2-3_char_cosine_0.6	0.041	Merck
0.277	BYFUSION_3-3_char_cosine_0.6	0.147	Feedback bis
0.266	BYFUSION_2-3_char_cosine_0.7	0.150	Feedback bis
0.265	BYFUSION_3-3_char_cosine_0.6	0.099	All In One
0.261	BYFUSION_3-3_char_cosine_0.6	0.150	Feedback
0.258	BYFUSION_3-3_char_cosine_0.7	0.157	Feedback
0.254	BYFUSION_2-3_char_cosine_0.6	0.102	Feedback
0.253	BYMAX_3-3_char_cosine	<b>0.160</b>	Feedback

## Résultats sur le jeu de test

Hamming	Paramètres	EMR	Corpus
<b>0.329</b>	BYFUSION_2-3_char_cosine_0.6	0.056	All In One
0.301	BYFUSION_2-3_char_cosine_0.6	0.078	Merck
0.270	BYFUSION_2-3_char_cosine_0.6	0.074	Feedback bis
0.269	BYFUSION_2-3_char_cosine_0.6	0.099	Feedback
0.249	BYFUSION_3-3_char_cosine_0.6	0.085	All In One
0.243	BYFUSION_2-3_char_cosine_0.7	0.090	All In One
0.239	BYFUSION_3-3_char_cosine_0.6	0.099	Feedback bis
0.237	BYFUSION_2-3_char_cosine_0.7	0.117	Feedback
0.231	BYMAX_2-3_char_cosine	<b>0.120</b>	Feedback
0.231	BYFUSION_2-3_char_cosine_0.9	<b>0.120</b>	Feedback

## Conclusion

---

- Système de question réponse fondé sur la recherche de similarités de phrases.

- Système de question réponse fondé sur la recherche de similarités de phrases.
- Plusieurs corpus de référence du domaine.

- Système de question réponse fondé sur la recherche de similarités de phrases.
- Plusieurs corpus de référence du domaine.
- Méthode **interprétable** !