



**HAL**  
open science

# Les écrits académiques en français à l'aide des corpus numériques

Cristelle Cavalla

► **To cite this version:**

Cristelle Cavalla. Les écrits académiques en français à l'aide des corpus numériques. Doctorat. Les écrits académiques en français à l'aide des corpus numériques, Université de Tunis, Tunisie. 2024. hal-04527489

**HAL Id: hal-04527489**

**<https://hal.science/hal-04527489>**

Submitted on 30 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# LES ÉCRITS ACADÉMIQUES EN FRANÇAIS À L'AIDE DES CORPUS NUMÉRIQUES

Cristelle CAVALLA

# UN ÉCRIT ACADÉMIQUE C'EST :

**Viser l'argumentation :**

Expliquer clairement sa position

**Viser l'objectivité :**

Atténuer sa trace

→ Être présent mais pas trop ;)



## Google Books Ngram Viewer

🔍 attaquer une question, étudier une question|



1800 - 2019 ▾

**French (2019) ▾**

Case-Insensitive

Smoothing ▾

## VÉRIFICATIONS...

Écrit-on plus souvent « attaquer une question » ou « étudier une question » ?

Écrivez-vous « dire l'hypothèse » ou « émettre l'hypothèse » ?

# PROGRAMME

Présentation des outils

Utilisation des corpus

Discussion

# Lexicoscope

## ScienQuest – Corpus Écrits scientifiques en français

Corpus Textes

Bienvenue sur ScienQuest

Démo

Démo

fr

Comparaison

**CNRTL** Centre National de Ressources Textuelles et Lexicales

**Ortolang**  
Outils et Ressources pour un Traitement Optimisé de la LANGue

Accueil | Portail lexical | Corpus | Lexiques | Dictionnaires | Métalexicographie | Outils | Contact

EST SHS

### Base LST SHS

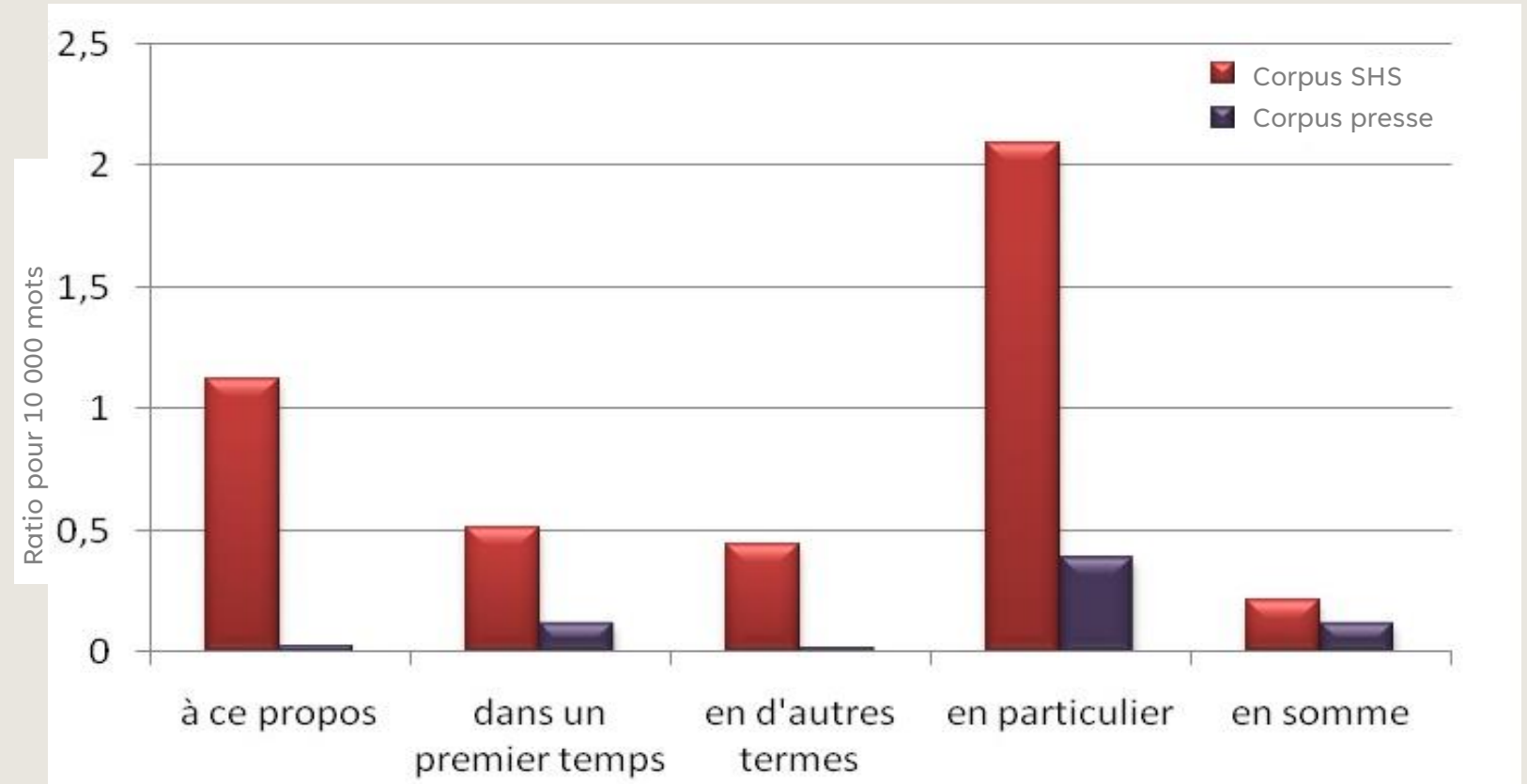
Lexique Scientifique Transdisciplinaire

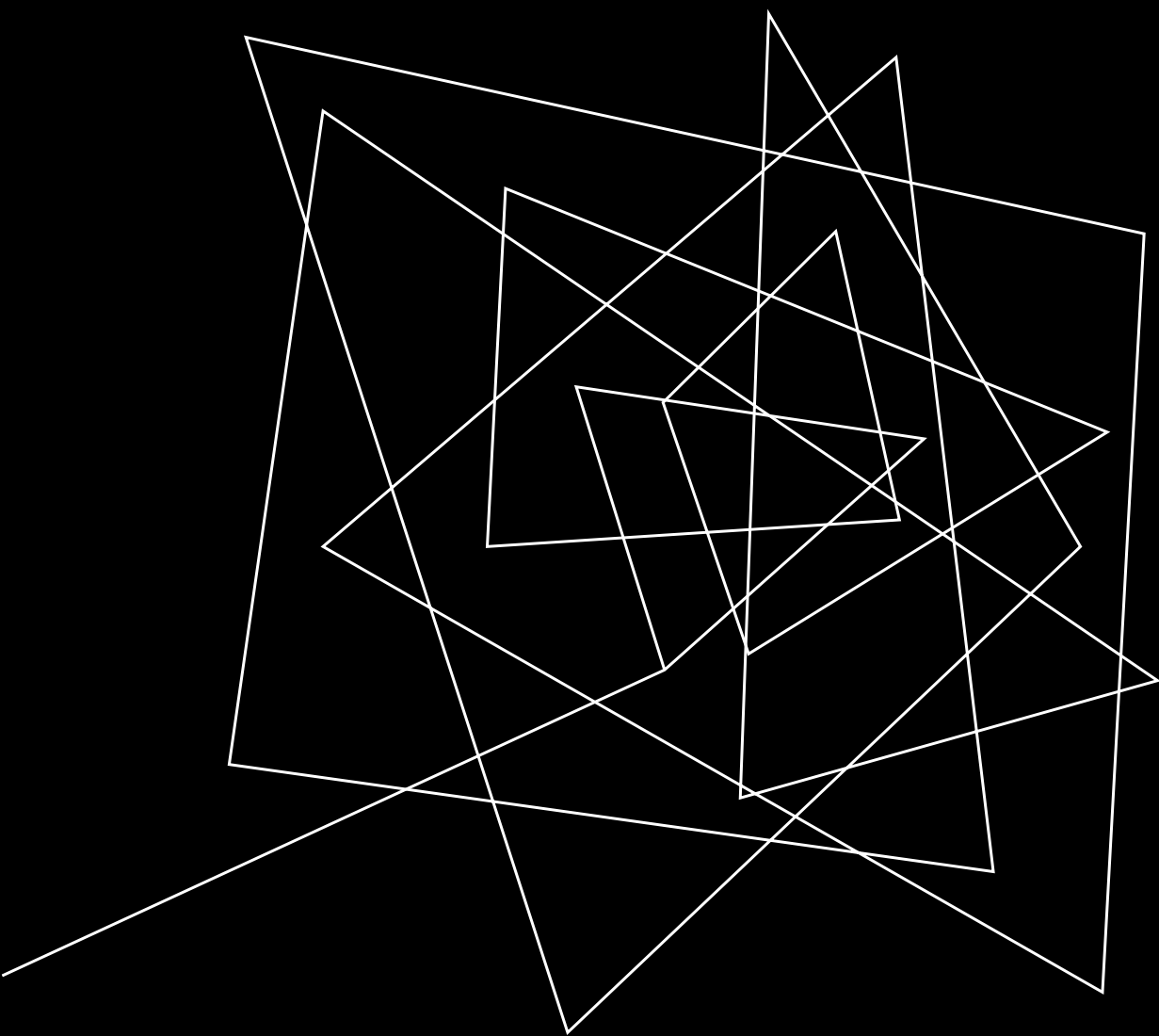
Rechercher dans les mots et expressions

<b>MOTS DU LST</b> Recherche des mots du Lexique Scientifique Transdisciplinaire	<b>EXPRESSIONS</b> Recherche des expressions scientifiques transdisciplinaires - voir les phrases typiques - voir les combinaisons lexicales - voir les locutions
<b>FONCTIONS</b> Présentation des fonctions rhétoriques du Lexique Scientifique Transdisciplinaire	<b>TRADUCTIONS</b> Recherche des mots du Lexique Scientifique Transdisciplinaire en 中文 (chinois)

# LES GENRES

- Corpus
  - Scientext : articles de Sciences Humaines et Sociales
  - Emolex : articles de presse
- Environ 2,5 millions de mots dans chaque corpus
  - **Spécificités de genres discursifs**





# CHERCHER « EN PARTICULIER »

Scienquest

Lexicoscope

LST



# « EN PARTICULIER » DANS SCIENQUEST

**ScienQuest – Corpus Écrits scientifiques en français**

Corpus | Textes | Recherche | Résultats | Outils | Connexion | À propos | Aide

Sémantique | Libre | Avancée

### Recherche

Vous pouvez effectuer des recherches de plusieurs manières: en mode "libre" (avec un assistant), ou en mode "avancé" (avec un langage de requête).

### Options de recherche

Max 1 000 000 résultats  
Contexte 10 mots (max 100)  
 Couper les contextes sur les limites de phrases.

### Recherche libre

En mode libre, vous pouvez

### Mot 1

Forme   
Lemme en particulier  
Catégorie

Chercher 913 occurrences

#	Contexte gauche	Occurrences	Contexte droit	Texte
1		En particulier	, les modèles statistiques sont d'un intérêt certain en	#1 - HDR - Sciences de l'éducation - Introduction
2	ostensifs à la portée des élèves est rendu très difficile à organiser ,	en particulier	l' utilisation du numérique explicite ( calculs ) est impossible	#1 - HDR - Sciences de l'éducation - Développement
3	conceptions relatives aux grandeurs ,	en particulier	le fait de faire un lien entre variation de l'	#1 - HDR - Sciences de l'éducation - Développement

<https://corpora.aiakide.net/scientext20/?do=SQ.setView&view=corpora>

# « EN PARTICULIER » DANS LEXICOSCOPE

Id : etudiant  
Mdp : etuCC2016

The screenshot shows the Lexicoscope web interface. At the top, there are navigation tabs: Statistiques, **Concordances**, Cooccurrences, Wordsketch, Partionnement des contextes, ALR, and Statistiques multivariées. There are also links for Déconnexion [cristelle], Données personnelles, and Aide. Below the tabs, there are toggle switches for KWIC/Déplié (checked) and ShowMetaData (unchecked). A navigation bar shows 'Global' selected and other time periods: 1950-59, 1960-69, 1970-79, 1980-89, 1990-99. The main content area is titled 'Concordances' and features a search bar with a magnifying glass icon and a search button. Below the search bar is a table with columns: Info, Sous-corpus, Contexte gauche, Nœud, and Contexte droit. The table contains seven rows of concordance data, each with a checkbox, an info icon, a list icon, a grid icon, a network icon, and a menu icon. The data rows are as follows:

Info	Sous-corpus	Contexte gauche	Nœud	Contexte droit
<input type="checkbox"/>	[1950]	d' autres réductions depuis lors ,	en	particulier au cours des récentes n
<input type="checkbox"/>	[1950]	Cela suppose	en	particulier que le syndicalisme accè
<input type="checkbox"/>	[1960]	vec nos partenaires atlantiques et	en	particulier avec les États-Unis .
<input type="checkbox"/>	[1970]		en	particulier aux pays francophones ,
<input type="checkbox"/>	[1980]	ont été expérimentées jusqu'ici ,	en	particulier aux États-Unis , ont mon
<input type="checkbox"/>	[1980]	" , qui survole son domaine	en	avion particulier et est fermement
<input type="checkbox"/>	[1990]	mporels , ce principe s' appliquant	en	particulier aux conflits territoriaux

[http://phraseotext.univ-grenoble-alpes.fr/lexicoscope\\_2.0/](http://phraseotext.univ-grenoble-alpes.fr/lexicoscope_2.0/)

# « EN PARTICULIER » DANS LE LST

<https://lst.demarre-shs.fr/>

Base LST SHS Ressources ▾ Langue ▾ Mentions légales

## en particulier

**Particulièrement ; surtout ; notamment.** (Wiktionnaire)  
Adverbe

Exemples -

Les auteurs tiennent à remercier les organisateurs de l'atelier consacré aux " Causes, formes et conséquences de la judiciaireisation ", qui leur ont permis de tester ainsi leurs premiers résultats, **en particulier** Christine Rothmayr, laquelle les a fait bénéficier de son excellente connaissance de la littérature internationale sur ce thème et de ses remarques judicieuses. [soc-art-35]

L'intensité des bouleversements de ces dernières décennies ne pourrait toutefois être que partiellement mise en évidence si la régionalisation du monde en trois grands ensembles n'était pas également prise en considération, **en particulier** l'intensité des bouleversements des rôles respectifs de l'Asie-Pacifique et de l'Amérique du Nord

La police nationale est omniprésente dans les médias, **en particulier** dans les séries télévisées(...). [infcom-art-5]

Sémantique +

Traductions +

# PROGRAMME

Présentation des outils

Utilisation des corpus

Discussion

# EXERCICE 1

L'analyse linguistique que nous proposons s'inscrit en réponse à une demande de la SNCF relative à la perception du confort global en train. Les objectifs du projet sont d'identifier les propriétés sémantiques du confort en train et leurs relations de dépendance, à partir d'analyses linguistiques et cognitives. Cet article porte sur la manière dont les formes linguistiques en contexte, utilisant les ressources de la langue mises en discours, renseignent sur les structures cognitives construites à partir des perceptions sensorielles. On tente ici d'identifier, à partir de l'analyse des formes adjectivales, les représentations individuelles et partagées qui se construisent dans les discours des voyageurs, lorsqu'on les interroge sur leur expérience sensible du confort. <...> La première question que l'on se pose alors est la suivante : le confort en train est-il une catégorie cognitive de ce type ?

1. Cet extrait se situe dans **quelle partie de l'article ?**

**Pourquoi ?**

A. résumé // B. introduction // C. théorie // D. méthodologie // E. conclusion

2. **Découpez l'extrait** en paragraphes . Dans le tableau suivant, relevez les **mots** qui introduisent le découpage :

Contexte	Sujet	Objectif	Question

## Contexte

- **L'analyse linguistique que nous proposons s'inscrit** en réponse à une demande de la SNCF relative à la perception du confort global en train. Les objectifs du projet sont d'identifier les propriétés sémantiques du confort en train et leurs relations de dépendance, à partir d'analyses linguistiques et cognitives.

## Sujet

- **Cet article porte sur** la manière dont les formes linguistiques en contexte, utilisant les ressources de la langue mises en discours, renseignent sur les structures cognitives construites à partir des perceptions sensorielles.

## Objectif

- **On tente ici d'identifier**, à partir de l'analyse des formes adjectivales, les représentations individuelles et partagées qui se construisent dans les discours des voyageurs, lorsqu'on les interroge sur leur expérience sensible du confort.

## Question de recherche

- **La première question que l'on se pose alors est la suivante** : le confort en train est-il une catégorie cognitive de ce type ?

# LISTE DES ÉLÉMENTS



Trouvez-en  
d'autres

- Présenter le sujet : → **Cet article porte sur**
- Introduire les questions pour cet article : → **La première question que l'on se pose alors est la suivante**
  - Présenter le contexte : → **L'analyse linguistique que nous proposons s'inscrit (en / dans)**
  - Présenter les objectifs de l'article : → **On tente ici de**



# CES ÉLÉMENTS SONT DES FORMES SPÉCIALES

Presque des expressions

Des formes presque figées

Se sont des routines que vous pouvez utiliser autant de fois que vous le souhaitez

On peut les utiliser de la sorte et on les trouve souvent dans les écrits universitaires

## Corpus Termith\_fr

Chercher le lemme « question »

## Concordances

Parcourez les exemples : la globalité du sous-corpus ; le sous-corpus en linguistique.

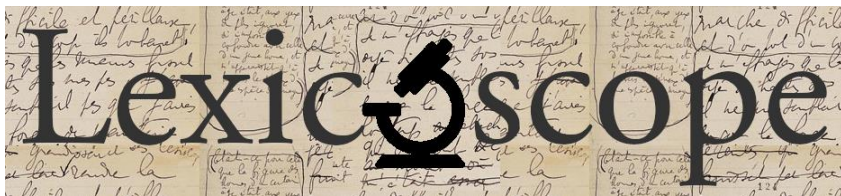
## Wordsketch

Dans la globalité du sous-corpus, quel est le verbe le plus fréquent dont « question » est objet ?

En *Linguistique*, quel est le 2<sup>e</sup> verbe le plus fréquent dont « question » est objet ?

## Cooccurrences

« Répondre » n'est pas très fréquent avec « question ». Cherchez des exemples pour voir à quel moment ou dans quelle discipline on utilise « répondre » avec « question ».



# PRÉSENTER DES HYPOTHÈSES



Consigne :

Soulignez les éléments linguistiques qui indiquent la/les hypothèse/s.

PRÉSENTER DES  
HYPOTHÈSES

Dans le domaine de l'analyse syntaxique, Gala (2003) fait l'hypothèse que plus une machine dispose d'une quantité importante de données, plus elle est en mesure de prendre les bonnes décisions. D'après l'auteur, il est donc nécessaire de disposer d'un corpus initial de grande taille afin d'augmenter le nombre d'échantillons utilisés pour calculer des probabilités de rattachement (Frérot, 2005).

Frérot, C. (2005). *Construction et évaluation en corpus variés de lexiques syntaxiques pour la résolution des ambiguïtés de rattachement prépositionnel*. Thèse de Doctorat, Linguistique et TAL. Université Stendhal, Grenoble.

Consigne :

Soulignez les éléments linguistiques qui indiquent la/les hypothèse/s.

PRÉSENTER DES  
HYPOTHÈSES

Dans le domaine de l'analyse syntaxique, Gala (2003) **fait l'hypothèse que** plus une machine dispose d'une quantité importante de données, plus elle est en mesure de prendre les bonnes décisions. D'après l'auteur, il est donc nécessaire de disposer d'un corpus initial de grande taille afin d'augmenter le nombre d'échantillons utilisés pour calculer des probabilités de rattachement (Frérot, 2005).

Frérot, C. (2005). *Construction et évaluation en corpus variés de lexiques syntaxiques pour la résolution des ambiguïtés de rattachement prépositionnel*. Thèse de Doctorat, Linguistique et TAL. Université Stendhal, Grenoble.

# ÉLÉMENTS POUR PRÉSENTER UNE HYPOTHÈSE

- Présenter une hypothèse : → Faire l'hypothèse que

Trouvez d'autres exemples avec ScienQuest

# SCIENQUEST : ÉCRITS SCIENTIFIQUES

## Entrer dans le corpus français

### Écrits scientifiques en français

Ce corpus a été conçu pour être représentatif des différents genres et disciplines scientifiques. [En savoir plus.](#)

Sélectionner  
205 textes  
6M mots  
Syntex

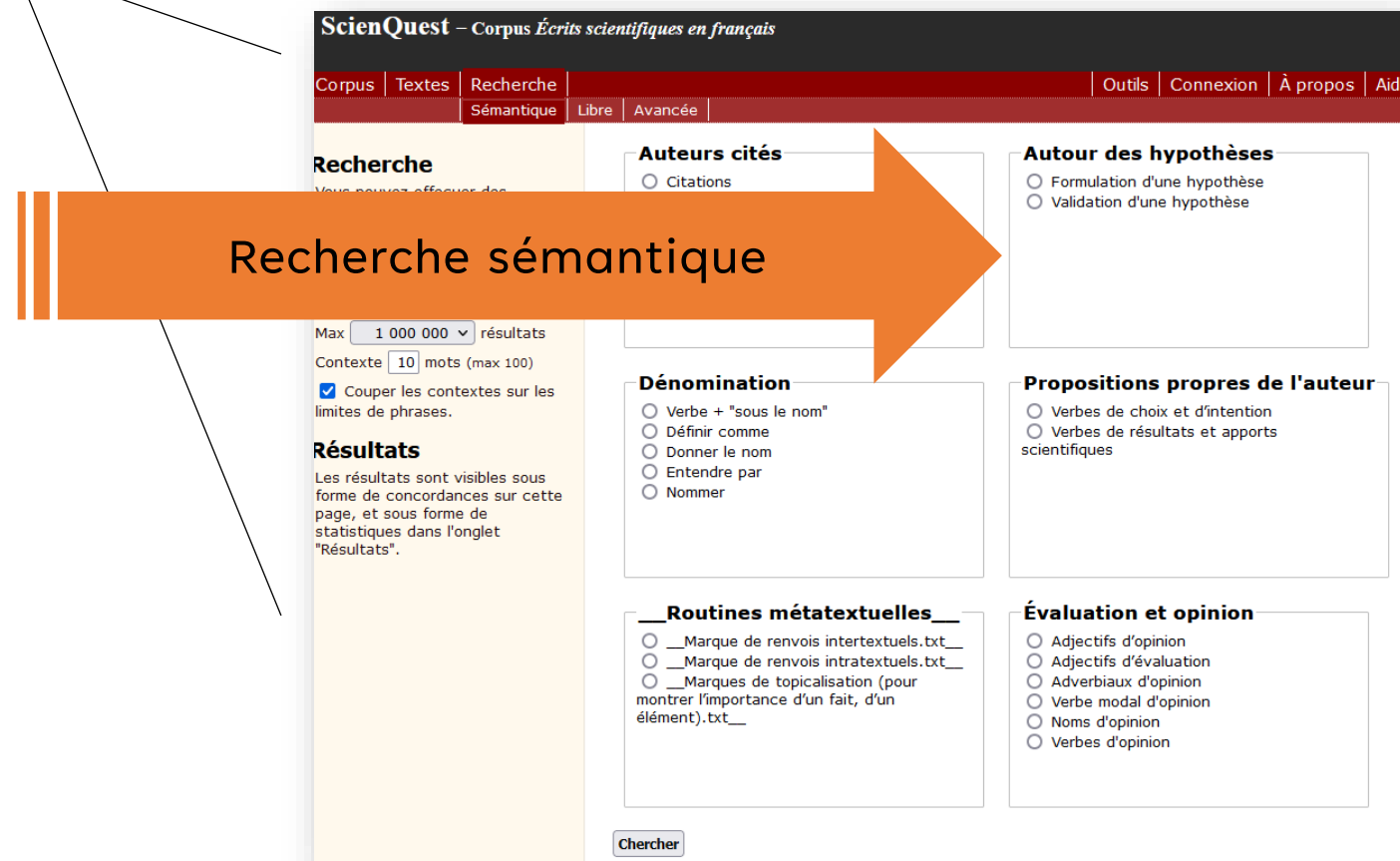
Choisissez des textes  
Ou pas ;)

## Trois types de recherches

The image displays three screenshots of the ScienQuest web application interface, illustrating different search options:

- Libre:** The top screenshot shows the 'Recherche' section with the 'Libre' mode selected. It features a search box for 'Mot 1' with fields for 'Forme', 'Lemme', and 'Catégorie', and a 'Chercher' button.
- Avancée:** The middle screenshot shows the 'Recherche' section with the 'Avancée' mode selected. It includes a 'Requête' field and a 'Chercher' button.
- Sémantique:** The bottom screenshot shows the 'Recherche' section with the 'Sémantique' mode selected. It features a 'Requête' field and a 'Chercher' button, along with various semantic search options like 'Auteurs cités', 'Autour des hypothèses', 'Dénomination', 'Propositions propres de l'auteur', 'Routines métatextuelles', and 'Évaluation et opinion'.

# SCIENQUEST : HYPOTHÈSE



**ScienQuest – Corpus Écrits scientifiques en français**

Corpus | Textes | Recherche | Outils | Connexion | À propos | Aide

Sémantique | Libre | Avancée

**Recherche**  
Vous pouvez afficher des

**Auteurs cités**  
 Citations

**Autour des hypothèses**  
 Formulation d'une hypothèse  
 Validation d'une hypothèse

**Dénomination**  
 Verbe + "sous le nom"  
 Définir comme  
 Donner le nom  
 Entendre par  
 Nommer

**Propositions propres de l'auteur**  
 Verbes de choix et d'intention  
 Verbes de résultats et apports scientifiques

**Routines métatextuelles**  
 \_\_Marque de renvois intertextuels.txt\_\_  
 \_\_Marque de renvois intratextuels.txt\_\_  
 \_\_Marques de topicalisation (pour montrer l'importance d'un fait, d'un élément).txt\_\_

**Évaluation et opinion**  
 Adjectifs d'opinion  
 Adjectifs d'évaluation  
 Adverbiaux d'opinion  
 Verbe modal d'opinion  
 Noms d'opinion  
 Verbes d'opinion

Max 1 000 000 résultats  
Contexte 10 mots (max 100)  
 Couper les contextes sur les limites de phrases.

**Résultats**  
Les résultats sont visibles sous forme de concordances sur cette page, et sous forme de statistiques dans l'onglet "Résultats".

Chercher

**Recherche sémantique**



# PRÉSENTER SON SUJET



Consigne:

A partir de l'extrait ci-dessous, il est possible de déterminer le sujet de l'article. Quel est-il ? Proposez des synonymes ou reformulations.

Le sujet

Dans cet article, nous nous intéresserons à l'évolution de l'activité sur la plate-forme pendant les deux premières années d'usage par les étudiants (2005 et 2006). Nous montrerons quels facteurs ont permis l'émergence d'une double dynamique : une dynamique d'écriture et une dynamique communautaire, les deux se trouvant dans une relation d'interdépendance réciproque. (Varga, 2007)

Varga, R. (2007). Évolution des usages d'une plate-forme collaborative, Conditions d'émergence d'une double dynamique communautaire et d'écriture. *Congrès International ARE (Actualité de la Recherche en Éducation et en Formation)*, Aug 2007, Strasbourg.

Consigne:

A partir de l'extrait ci-dessous, il est possible de déterminer le sujet de l'article. Quel est-il ? Proposez des synonymes ou reformulations.

Le sujet

Dans cet article, nous nous intéresserons à l'évolution de l'activité sur la plate-forme pendant les deux premières années d'usage par les étudiants (2005 et 2006). Nous montrerons quels facteurs ont permis l'émergence d'une double dynamique : une dynamique d'écriture et une dynamique communautaire, les deux se trouvant dans une relation d'interdépendance réciproque. (Varga, 2007)

# ELÉMENTS POUR PRÉSENTER LE SUJET

- Présenter le sujet :
  - Nous nous intéresserons à
  - Nous montrerons que

Trouvez d'autres exemples avec le LST

# LST : LEXIQUE SCIENTIFIQUE TRANSDISCIPLINAIRE

**Base LST SHS**  
Lexique Scientifique Transdisciplinaire

Rechercher dans les mots et expressions

**MOTS DU LST**  
Recherche des mots du Lexique Scientifique Transdisciplinaire

**EXPRESSIONS**  
Recherche des expressions scientifiques transdisciplinaires

- Voir les phrases typiques
- Voir les combinaisons lexicales
- Voir les locutions

**FONCTIONS**  
Présentation des fonctions rhétoriques du Lexique Scientifique Transdisciplinaire

**TRADUCTIONS**  
Recherche des mots du Lexique Scientifique Transdisciplinaire en  
中文 (chinois)

**Classes sémantiques**  
Accès à la classification

**Textes annotés**  
Visualisation d'articles

**À propos**  
Page d'informations

# LST : FONCTION « SUJET »

## aborder un sujet

Mot principal	<a href="#">sujet_N</a>
Mot associé	<a href="#">aborder_V</a>
Forme de base	aborder un sujet
Patron syntaxique	Complément d'objet
Fonction lexicale	Oper1?
Fonctions	<a href="#">Définir thème essentiel</a> ? - <a href="#">Sujet</a> ?

### Exemples

Avant d' **aborder ce sujet** dans les contributions suivantes, et même si le modèle des revues ne peut être entièrement transposé à ce nouvel objet, il est utile de tirer les enseignements de la révolution issue du développement des ressources numériques dans les services documentaires du monde universitaire et de la recherche. [infcom-art-232]

### Statistiques

Nombre d'occurrences	43
Répartition (sur 100 tranches)	TF
Répartition (sur 10 disciplines)	6
Force de l'association	308

# ELÉMENTS POUR PRÉSENTER LE SUJET

Présenter le sujet de la recherche :

→ Cet article porte sur

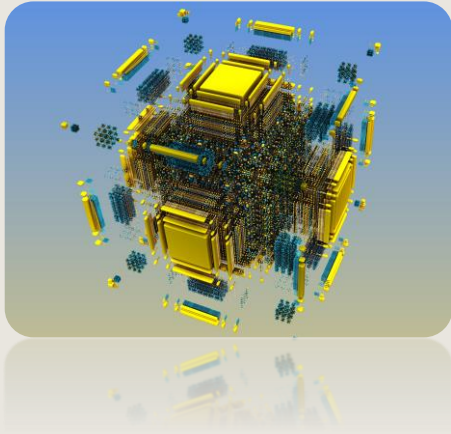
→ Dans cet article nous nous intéresserons à

→ Nous montrerons que

→ Nous nous intéresserons ici à

# POSER DES QUESTIONS





# CHERCHEZ DES EXEMPLES DANS LES CORPUS

Poser des questions

# PROGRAMME

Présentation des outils

Utilisation des corpus

Discussion



MERCI

LES ÉCRITS ACADÉMIQUES EN FRANÇAIS À  
L'AIDE DES CORPUS NUMÉRIQUES

**Cristelle CAVALLA**

[Cristelle.Cavalla@sorbonne-nouvelle.fr](mailto:Cristelle.Cavalla@sorbonne-nouvelle.fr)

**Sorbonne  
Nouvelle**  DILTEC - EA 2288  
Didactique des langues,  
des textes et des cultures

1<sup>er</sup> mars 2024