



# Aide à l'écriture scientifique : une entrée outillée

**Cristelle CAVALLA**



## Présentation



1. **Hypothèse et objectifs**
2. **Outils d'analyses et pédagogiques**
  - Outil pour séquence
  - Type de corpus
  - Analyses linguistiques
  - Approche pédagogique
3. **Séquence didactique**
  - Contexte d'enseignement
  - Introduction des corpus
  - Introduction de la phraséologie
4. **Conclusion**

# Objectifs généraux et hypothèse

## Objectifs linguistiques

- ▶ Linguistique de corpus
- ▶ Genre « écrit académique »



## Objectifs didactiques

- ▶ Utilisation du corpus en classe de langue
- ▶ Appropriation des spécificités des écrits scientifiques

Discours académique

- Normes langagières du genre
- Normes professionnelles

**Hypothèse didactique (affect) :**  
L'entrée par le corpus numérique aide à aborder la langue de façon sereine

# Objectifs pour l'apprenant

Être capable  
d'utiliser des  
corpus  
numériques pour  
rédiger



Être capable  
d'utiliser les  
éléments  
linguistiques  
appropriés à cet  
écrit



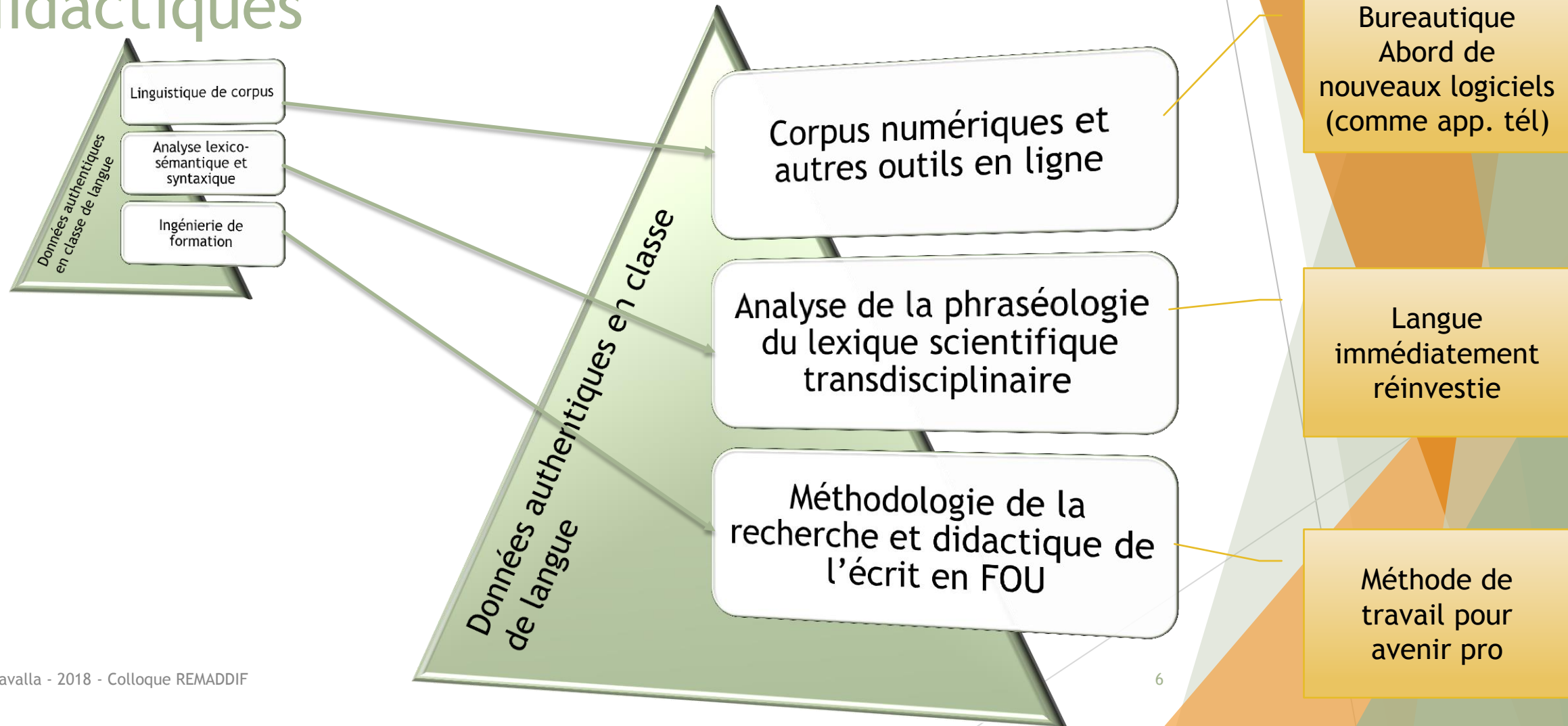
Être capable  
de rédiger un  
écrit  
académique

## Présentation



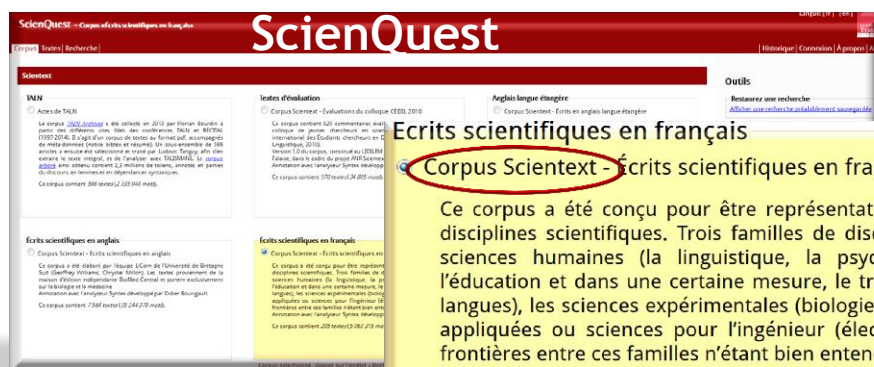
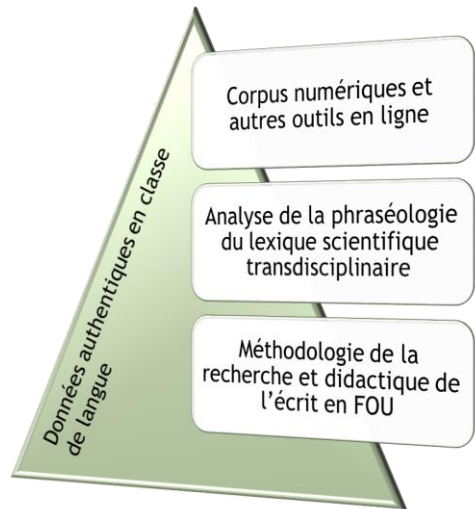
1. **Hypothèse et objectifs**
2. **Outils d'analyses et pédagogiques**
  - Outils pour séquence
  - Type de corpus
  - Analyses linguistiques
  - Approche pédagogique
3. **Séquence didactique**
  - Contexte d'enseignement
  - Introduction des corpus
  - Introduction de la phraséologie
4. **Conclusion**

# Outils pour l'élaboration de séquences didactiques





# Corpus utilisés : le « déjà-là » des étudiants



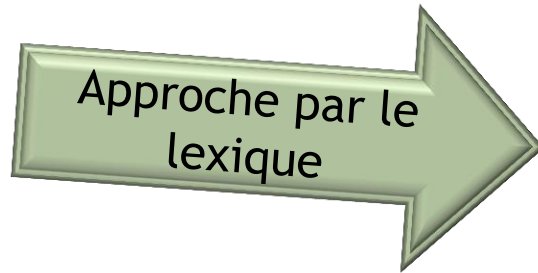
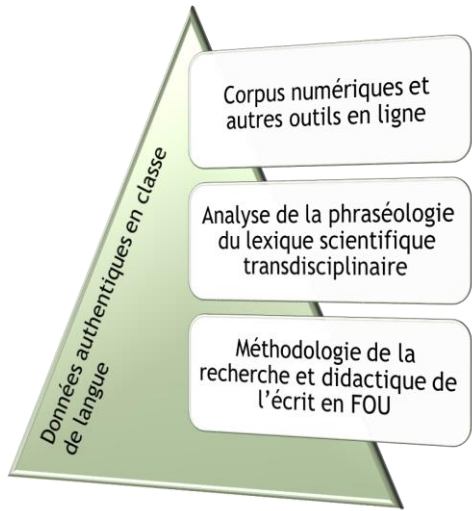
**Écrits scientifiques en français**

**Corpus Scientext - écrits scientifiques en français**

Ce corpus a été conçu pour être représentatif des différents genres et disciplines scientifiques. Trois familles de disciplines sont incluses : les sciences humaines (la linguistique, la psychologie, les sciences de l'éducation et dans une certaine mesure, le traitement automatique des langues), les sciences expérimentales (biologie, médecine) et les sciences appliquées ou sciences pour l'ingénieur (électronique, mécanique), les frontières entre ces familles n'étant bien entendu pas étanches. Annotation avec l'analyseur Syntex développé par Didier Bourgauf.

Ce corpus contient 205 textes (5 063 315 mots).

# Analyses linguistiques utilisés

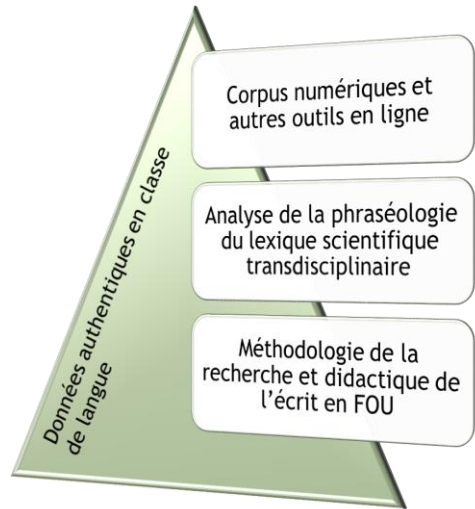


## Étude de la **phraséologie transdisciplinaire**

- Définition (Tutin et Pecman)
  - Extraction
- Analyses sémantique et syntaxique
- Approche Sens-Texte (Mel'cūk, Polguère)



# Approche pédagogique utilisée



Objectif : autonomie

■ Directe : leurs questions

- comment énoncer ses questions de recherche / ses hypothèses ? Comment présenter son sujet ?...

■ Approche inductive : leurs besoins

- Prise de conscience du contenu de l'écrit académique
- Organisation du contenu
- Recherche des éléments pour énoncer le contenu

■ Onomasiologie : vers autonomie

- Entrées sémantiques

Approche directe

Objectif : spécifier un élément linguistique

■ Indirecte : arrêt sur image

- concordancier sur un seul élément ou sur une classe sémantique

■ Comprendre l'utilisation : autonomie de la méthode d'interrogation épistémologique (comment fonctionne cette langue ?)

Approche indirecte

## Présentation



1. **Hypothèse et objectifs**
2. **Outils d'analyses et pédagogiques**
  - Outils pour séquence
  - Type de corpus
  - Analyses linguistiques
  - Approche pédagogique
3. **Séquence didactique**
  - Contexte d'enseignement
  - Introduction des corpus
  - Introduction de la phraséologie
4. **Conclusion**

# Public - Formation

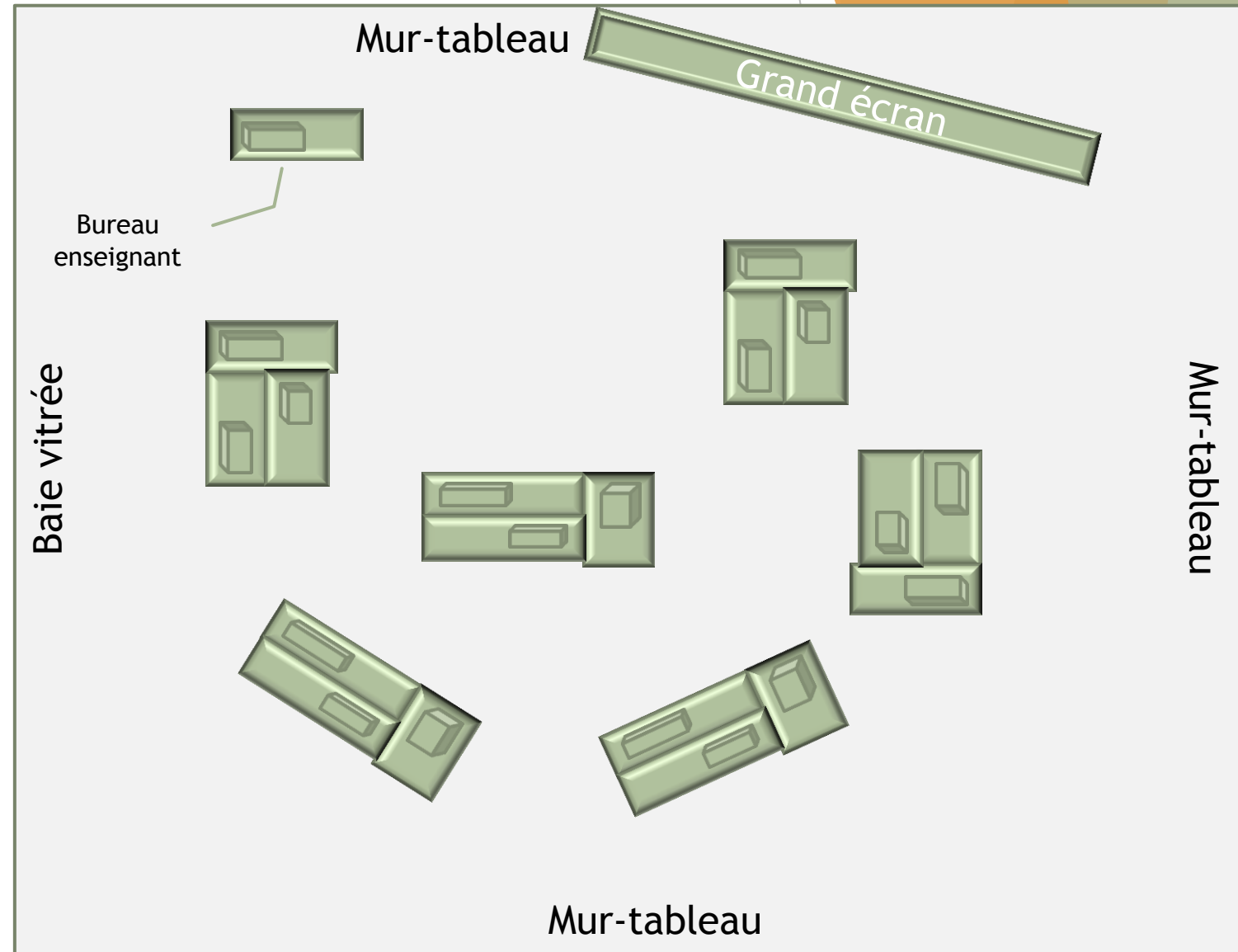
- ▶ Futurs journalistes, traducteurs, enseignants, juristes dans le domaine culturel, médiateur interculturel...

Master 2<sup>e</sup> année - Niveau B2-C1 - Mémoire à rédiger en français

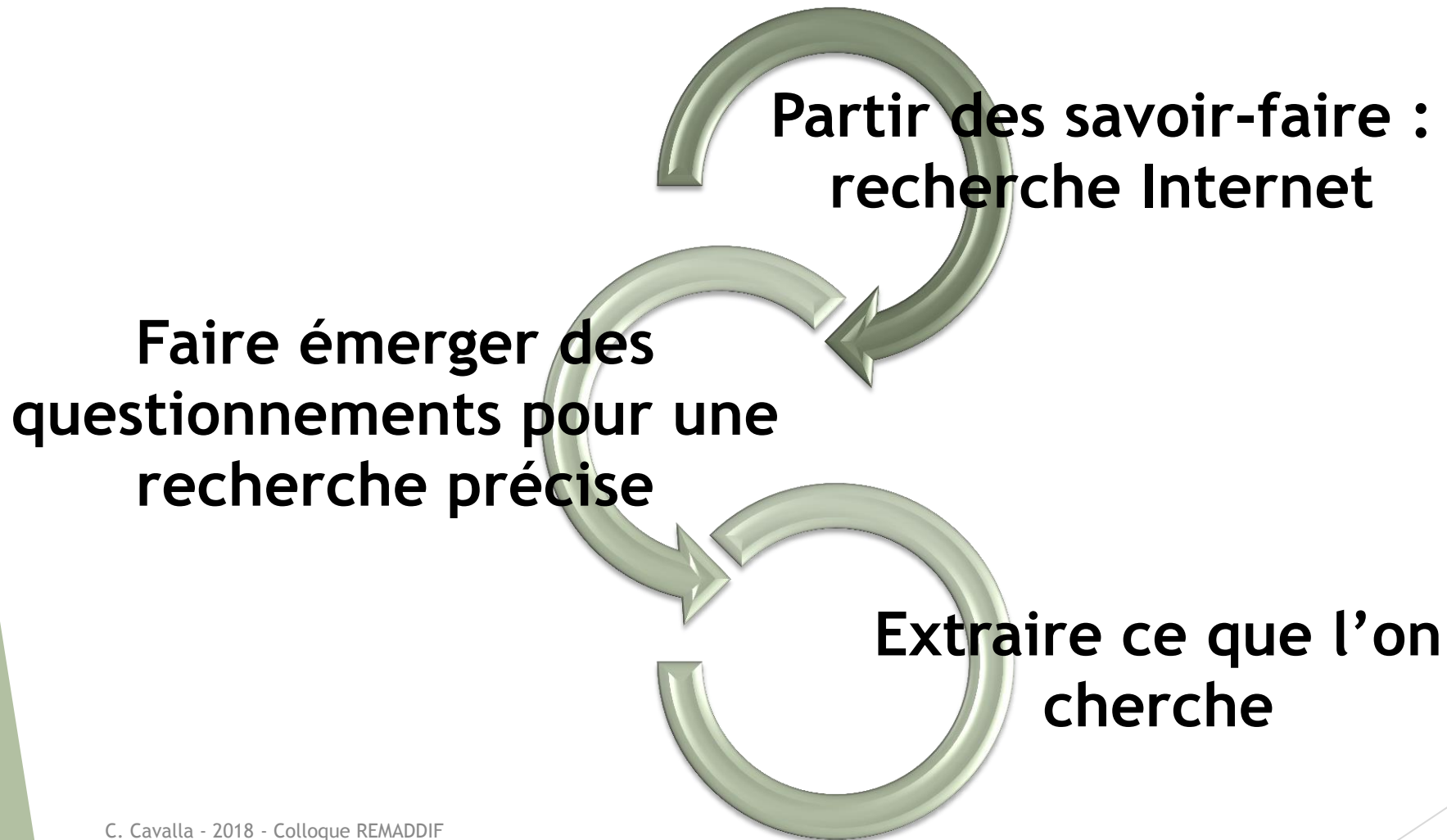


# Conditions d'enseignement

- ▶ Salle équipée
- ▶ Chacun un ordinateur
- ▶ En groupe rassemblés: en îlots
  - ▶ Échanges
  - ▶ Travail collaboratif



# Comment faire découvrir ?





# Approche Directe : recherche sur outils connus

Linguee

Google Books Ngram Viewer

Dans quel type de phrase trouve-t-on l'expression  
« en particulier » ?



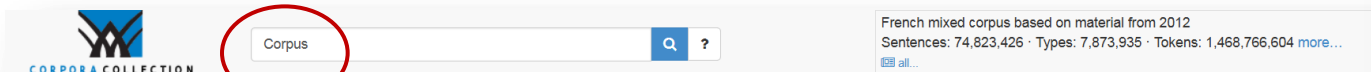
Comment sélectionner les exemples trouvés dans Google ?  
Comment vérifier les contextes ?

...

L'expression « en particulier » est utilisée dans quel contexte et par qui ?

# Approche Directe

## 1<sup>er</sup> corpus : des chiffres



Word: **Corpus** Number of occurrences: 1,221 Rank: 50,231 Frequency class: 16

See also: corpus, CORPUS

Similarity based on Cooccurrences: Corpus Christi | Christi

**Christi : 5 196 occurrences**

Examples: S'inscrivant dans le cadre du festival Corpus,...

Cooccurrences:

Corpus Christi (6,165), Christi (5,196), Domini (1,353), Texas (864), Habeas (798), Hermeticum (556), Corpus hippocratique (486), l'Habeas (486) College (309), Juris (299), Juris (299), Christianorum (287), Corpus Delicti (287), Corpus Christianorum (287), Delicti (287), ) (279), Inscriptionum ( Corpus inscriptionum latinorum (199), Ave Verum Corpus (199), Galveston (197), latinorum (192), solennité (182), procession (182), Jérôme Prieu (153), Justinien (152), Ave (141), Cambridge (133), d'Habeas (121), Langage (120), Rhésus (117), : (117), Andrea Bocelli (116), Prieur (112), Bocc (109), College (102), Medicorum (98), Etampois (97), langue (96), Linguistiques (95), Latinarum (88), Separatum (88)



Word: **corpus** Number of occurrences: 6 108 Rank: 17 112 Frequency class: 14

**textes : 2863 occurrences**

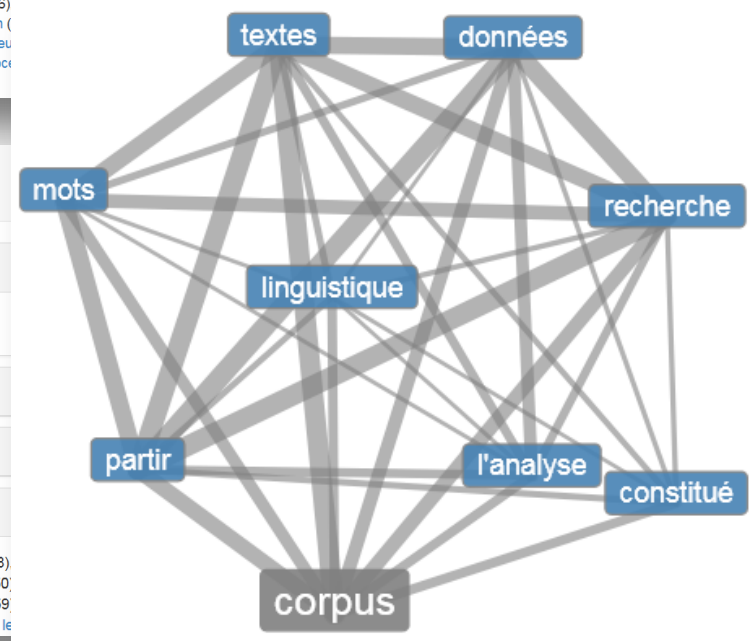
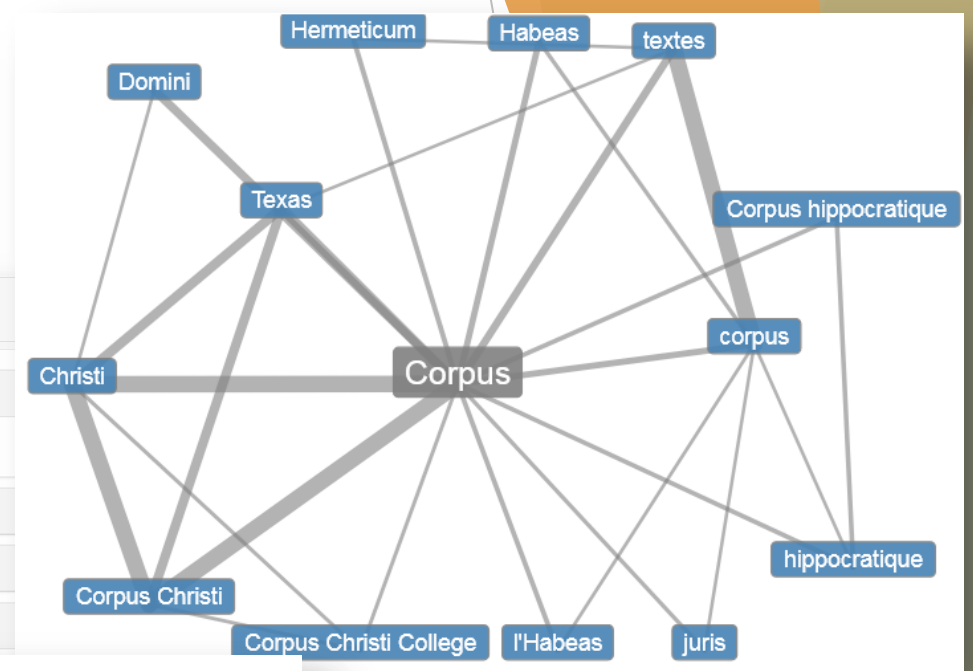
See also: Corpus, CORPUS

Similarity based on Cooccurrences: lexique | recueil | vocabulaire | catalogue | répertoire

Examples: Le corpus documentaire réuni pour cette étude...

Cooccurrences:

textes (2,863), l'habeas (1,539), habeas (1,294), d'habeas (941), partir (707), de (702), linguistique (675), mots (651), un (648), d'un (648) idéologique (448), connaissances (434), théorique (426), Corpus (418), linguistiques (411), oraux (398), et (390), ( (357), recherches (350) (309), constituer (306), documents (300), littéraires (289), outils (280), textuels (266), dans (266), sur (265), langues (260), d'analyse (259) littérature (233), T-LAB (231), législatif (230), automatique (229), documentaire (226), l'ensemble (225), différents (220), L'habeas (220), le



Découverte des extractions : Comment les lire ?

# Approche Directe : 2<sup>e</sup> corpus : un concordancier

Lemma : considérer

Observer à droite : comme / que

French mixed corpus based on material from 2012  
Sentences: 74.823.426 · Types: 7.873.935 · Tokens: 1.468.766.604 more...

Word: **considérer** Number of occurrences: 31,240 Rank: 4,600 Frequency class: 11

See also: Considérer, considérer, considérer, considérer, Considérer, CONSIDÉRER, considÉrer

Similarity based on Cooccurrences: penser | admettre | envisager | comprendre | imaginer

Examples: A défaut d'une meilleure estimation, l'Autorité a...

Concurrences:

comme (55,123), peut (18,457), que (12,163), faut (5,844), On (5,707), on (5,519), ne (3,181), l'on (2,655), à (2,415), un (2,356), qu'il (2,276), pas (2,038), une (1,965), il (1,681), de (1,555), donc (1,520), les (1,505), tendance (1,320), est (1,260), ce (1,196), pouvons (1,097), qu'on (1,086), non (992), étant (964), il (952), nous (939), ces (910), doit (891), consiste (716), serait (707), devons (699), le (699), plutôt (656), si (652), pourrait (633), cette (622), façon (607), (598), (592), la (570), point (570), refuse (566), ou (552), entière (536), s'agit (530), manière (522), mais (518), s'accordent (514), convient (510), qu'un (502), qu'une (485), n'est (479), peut-on (451), cela (443), faudrait (441), amène (440), même (440), cas (429), suffit (415), c'est (413)

## Compleat Lexical Tutor

Home > Concordancers > French Input [«Back】 (Back keeps original settings) Copiable extract-Link to this data >> here

Concordance for lemma *considérer* in Fr\_le\_monde.txt sorted 1 wd left of key [Dictionnaire] [Fren\_Eng] [Speak] [Fr-f

Extract: [All] [0] [any] [10] [20] [30] [50] Go >

[lemma] [considérer] Le Monde (1998) 1 110 392 [sorted] 1 wd [left] [+assoc] [on left]

10. [ ] tre forme de procès eût été le pire des **affronts**. [CONSIDÉRANT](#) que la meilleure défense était l'attaq

11. [ ] -Il poursuivi "Le Pakistan, a-t-il enfin **ajouté**, [CONSIDÈRE](#) que ces deux points doivent être discuté

12. [ ] CDU et le SPD, un scénario que 40 % des **Allemands** [CONSIDÈRENT](#) comme le plus probable à huit mois des

13. [ ] e tribunal administratif. Les juges avaient **alors** [CONSIDÉRÉ](#) que l'on ne pouvait fixer "de conditions

14. [ ] c à Alcatel à amplifier son rebond. Les **analystes** [CONSIDÈRENT](#), de façon générale, que le groupe en a

15. [ ] ssi lourdes, reconnaît Compaq. Certains **analystes** [CONSIDÈRENT](#) que la remise en confiance de la clien

16. [ ] ux seules ventes de logiciels. Certains **analystes** [CONSIDÈRENT](#) qu'en développant une activité de cons

17. [ ] au Collège de France, âgé de soixante-quinze **ans**, [CONSIDÉRÉ](#) comme l'un des plus grands économistes f

18. [ ] "affaires générales" d'Elf-Aquitaine, **aujourd'hui** [CONSIDÉRÉ](#) comme le personnage central de l'affaire

19. [ ] , Charlotte Perriand créa des meubles **aujourd'hui** [CONSIDÉRÉS](#) comme des classiques de la modernité. E

20. [ ] dent le relèvement ou la grâce, et quinze **autres**, [CONSIDÉRÉES](#) comme ne remplissant pas les critères

21. [ ] bunal impartial", les juges de Strasbourg **avaient** [CONSIDÉRÉ](#) que la cour d'assises n'avait pas procé

22. [ ] uption, la haute juridiction administrative **avait** [CONSIDÉRÉ](#) que le contrat conclu par le maire avec

23. [ ] de l'ancien ministre GATTEGNO HERVE **Un** [avocate](#), [CONSIDÉRÉE](#) comme une amie proche de Roland Dumas,

24. [ ] ri Philippi, Louis Schweitzer et Patrick **Baudry**), [CONSIDÈRE](#) que l'information judiciaire ne peut étr

25. [ ] es moins "rébarbatives", comme en Staps. M. **Borel** [CONSIDÈRE](#) enfin que les sciences sont actuellement

26. [ ] rovenance de Londres. Le gouvernement **britannique** [CONSIDÈRE](#) comme superfétatoire l'adoption d'une lo

27. [ ] France n'est pas davantage pour la **centralisation** [CONSIDÉRÉE](#) comme une sorte de dogme. Quant à l'All

28. [ ] oire le fait qu'il y a moins de cent ans la **Chine** [CONSIDÉRAIT](#) comme relevant de sa souveraineté la p

29. [ ] e l'écrivain silencieux". La thèse de Marc **Comina** [CONSIDÈRE](#) les textes de l'écrivain et ceux des cri

30. [ ] se trouve dans une situation juridique **complexe**, [CONSIDÉRÉE](#) par Me Schoenbach comme une "injustice

31. [ ] à éviter, comme les grands conglomérats **coréens**. [CONSIDÉRÉS](#) à tort comme une source de réussite, il

32. [ ] n, le président de l'Assemblée nationale **cubaine**, [CONSIDÉRÉ](#) comme un partisan de l'ouverture, tandis

33. [ ] arck un précurseur des nazis. Hitler, **d'ailleurs**, [CONSIDÉRAIT](#) que le chancelier de Guillaume II, mal

34. [ ] : que l'Europe de demain cesse tout d'un coup **de** [CONSIDÉRER](#) les pays d'Asie comme un modèle de réf

035. [ ] du Rouergue"; les plus modestes se contentent **de** [CONSIDÉRER](#) le lieu comme un écrin de culture et de

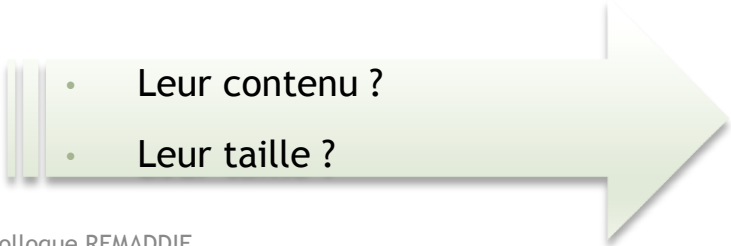
036. [ ] PIB européen), il paraît, en effet, difficile **de** [CONSIDÉRER](#) que l'UEM puisse être, de ce point de v

037. [ ] 'accusant de diviser le peuple et de continuer **de** [CONSIDÉRER](#) comme non israélienne la moitié de la p

# Approche directe : 1<sup>re</sup> réflexion sur l'outil

## Contextualisation

Comparer les sorties entre sites :  
Leur présentation ? Quel type  
d'information ?...

- 
- Leur contenu ?
  - Leur taille ?

- Textes différents donc langue différente
- Taille différente donc résultats à harmoniser
- Année des textes...

# Approche directe : prise en main personnelle

Comparer les deux corpus suivants.  
Quelles sont vos conclusions ?

- FR EN

## Lexicoscope

- Exploration des profils combinatoires -

Sélection du Corpus Requête Paramètres Sessions sauvegardées Guide

Corpus monolingue   
Corpus parallèle

Langue fr

- Corpus littéraire du projet phraséotexte - Policier
- Corpus littéraire du projet phraséotexte - Sentimental
- Corpus littéraire du projet phraséotexte - Historique
- Corpus littéraire du projet phraséotexte - Science fiction
- Corpus journalistique français**
- Corpus littéraire français
- Corpus journalistique français (emoBase - étiquettes Connexor)
- Corpus littéraire français (emoBase - étiquettes Connexor)
- Corpus journalistique français (emoBase - étiquettes Connexor - 20M)
- Corpus journalistique français (emoBase - étiquettes Connexor - 16M)

**Description**

Corpus sélectionné : nombre de mots = **112 280 979** ( **299 125** textes )

## ScienQuest – Corpus «Écrits scientifiques en français»

Corpus Textes Recherche Résultats

# ScienQuest

### Scientext

**TALN**

Actes de TALN

Le corpus [TALN Archives](#) a été collecté en 2013 par Florian Bourdin à partir des différents sites Web des conférences TALN et RÉCITAL (1997-2014). Il s'agit d'un corpus de textes au format pdf, accompagnés de méta-données (notice bibtex et résumé). Un sous-ensemble de 586 articles a ensuite été sélectionné et traité par Ludovic Tanguy, afin d'en extraire le texte intégral, et de l'analyser avec TALISMANE. Le [corpus arboré](#) ainsi obtenu contient 2,3 millions de tokens, annotés en parties du discours, en lemmes et en dépendances syntaxiques.

Ce corpus contient *586 textes (2 335 943 mots)*.

**Textes d'évaluation**

Corpus Scientext - Évaluations du colloque CÉDIL 2010

Ce corpus contient 520 commentaires évaluatifs de relecteurs pour un colloque de jeunes chercheurs en sciences du langage (Colloque international des Étudiants chercheurs en Didactique des Langues et en Linguistique, 2010).  
Version 1.0 du corpus, constitué au LIDILEM par Françoise Boch et Achille Falaise, dans le cadre du projet ANR Scientext.  
Annotation avec l'analyseur Syntex développé par Didier Bourigault.

Ce corpus contient *570 textes (34 805 mots)*.

**Anglais langue étrangère**

Corpus Scientext - Écrits en anglais langue étrangère

Ce corpus comporte des travaux d'apprenants universitaires français écrivant en anglais, principalement des étudiants de 2e et 3e année du cursus d'anglicistes apprenant à rédiger de textes argumentatifs longs (4500 mots) qui s'appuient sur des recherches documentaires approfondies.  
Version 1.0 du corpus, constitué au LLS par John Osborne, Alice Henderson et Robert Barr, dans le cadre du projet ANR Scientext.  
Annotation avec l'analyseur Syntex développé par Didier Bourigault.

Ce corpus contient *272 textes (1 020 146 mots)*.

**Écrits scientifiques en anglais**

Corpus Scientext - Écrits scientifiques en anglais

Ce corpus a été élaboré par l'équipe LICorn de l'Université de Bretagne Sud (Geoffrey Williams, Chrystel Millon). Les textes proviennent de la maison d'édition indépendante BioMed Central et portent exclusivement sur la biologie et la médecine  
Annotation avec l'analyseur Syntex développé par Didier Bourigault.

Ce corpus contient *7 564 textes (35 244 378 mots)*.

18



# Introduction de la phraséologie

- ▶ Pour chaque outil : recherche de mot isolé et de phrasème
  - ▶ Ngram Viewer : *mettre la table / dresser la table*
  - ▶ Corpus français de Leipzig : pas possible
  - ▶ Lextutor : *cause / remettre en cause*
  - ▶ Lexicoscope : *cap / passer le cap*
  - ▶ ScienQuest : *cap / passer le cap / émettre une hypothèse*
    - ▶ Entrée libre

**Lexicoscope**  
- Exploration des profils combinatoires -

Sélection du Corpus Requête Paramètres Sessions sauvegardées Guide

Concordances et profils combinatoires (cooccurrences)

Requête libre Requête avancée Requête multi-pivots

passer le cap

Concordances  
Cooccurrences

p.ex. "considération" ou "prendre en considération"

**ScienQuest** – Corpus «Écrits scientifiques en français»

Corpus Textes Recherche Résultats

Sémantique Libre Avancée

Mot 1  
Forme  
Lemme passer  
Catégorie

Mot 2  
Forme  
Lemme cap  
Catégorie

Relations syntaxiques  
Mot 2 - objet direct de (OBJ) Mot 1

Chercher 1 occurrence

#	Contexte gauche	Occurrences	Contexte
1	de l' enseignement , et qui n' arriveraient pas à	passer le cap	de la conceptualisation mathématique nous

Résultats de 1 à 1 sur un total

# Ils découvrent alors la recherche sémantique

Langue: [ fr ] [ en ]

ScienQuest – Corpus «Écrits scientifiques en français»

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

Sémantique | Libre | Avancée

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

# S'interrogent sur leurs besoins

- ▶ Réflexion sur l'écrit :  
quelles autres questions ?
  - ▶ Comment écrire mes questions de recherche?
  - ▶ Comment citer un auteur ?
  - ▶ ...

# Vers l'écriture

## Répondez aux questions suivantes

1

1. D'après vous, cet extrait vient de quelle partie de l'article ? Pourquoi ?  
A. résumé // B. introduction // C. théorie // D. méthodologie // E. conclusion
2. Découpez l'extrait en paragraphes et justifiez votre choix en précisant les mots et les thématiques de votre découpage à l'aide du tableau suivant :

Contexte	Sujet	Objectif	Question

3. Trouvez des expressions synonymes pour dire « on tente de... » à l'aide de l'outil CNTRL en ligne : <http://www.cnrtl.fr/synonymie/>

Précisions ensuite sur introduction / questions de recherche / citer un auteur...

2

## Découpage

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

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

46

## Liste des éléments

3

- 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

Trouvez-en d'autres

## Présentation



1. **Hypothèse et objectifs**
2. **Outils d'analyses et pédagogiques**
  - Outils pour séquence
  - Type de corpus
  - Analyses linguistiques
  - Approche pédagogique
3. **Séquence didactique**
  - Contexte d'enseignement
  - Introduction des corpus
  - Introduction de la phraséologie
4. **Conclusion**



# Démarche FOS

Étudiants de Master pour rédiger leurs mémoires

1/ Commande → analyse de la demande

Manque d'éléments spécifiques aux écrits académiques

2/ Analyse des besoins / premières hypothèses

3/ Contact avec les acteurs du terrain / collecte des données / confirmation ou infirmation des hypothèses

Attentes universitaires  
Création de corpus des genres d'écrits universitaires

4/ Traitement des données / analyse des discours

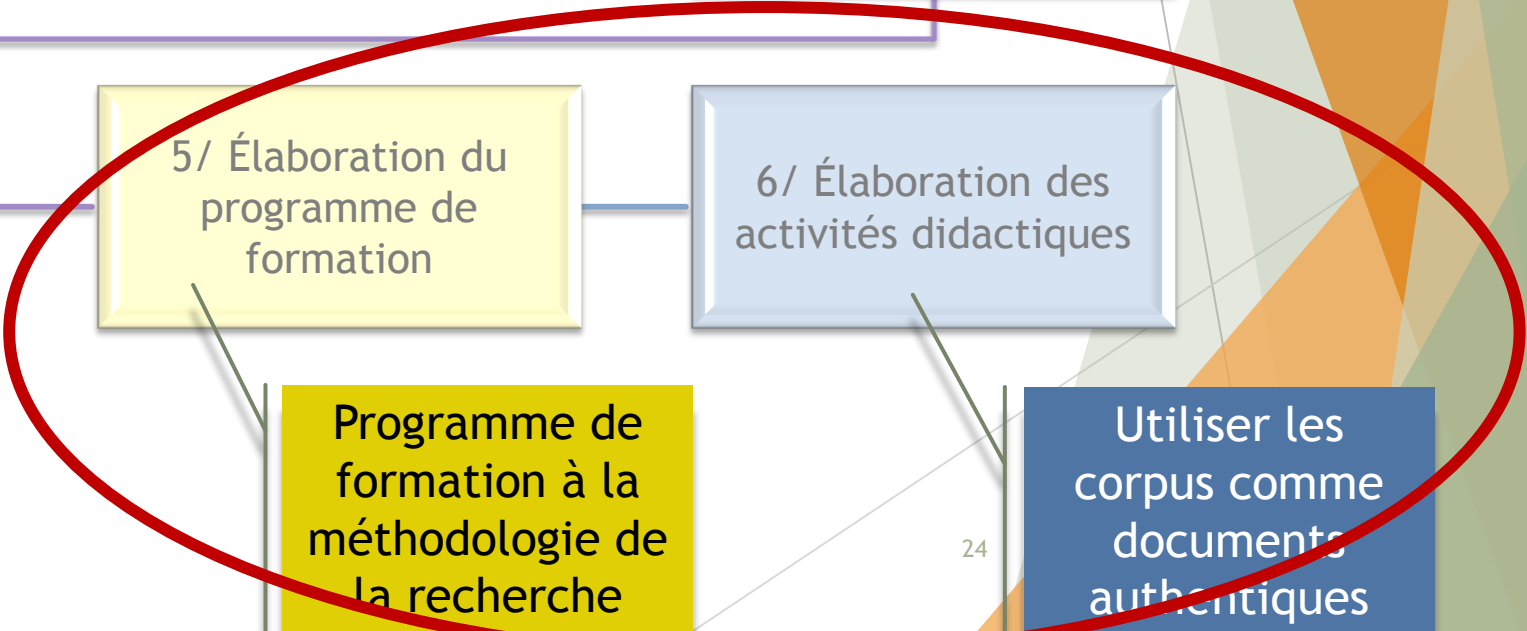
Analyse de la phraséologie transdisciplinaire

5/ Élaboration du programme de formation

Programme de formation à la méthodologie de la recherche

6/ Élaboration des activités didactiques

Utiliser les corpus comme documents authentiques



# Aspects affectifs

## **Hypothèse didactique (affect) :**

L'entrée par le corpus numérique aide à aborder la langue de façon sereine

Vérification

sur le terrain  
grâce aux  
témoignages des  
étudiants

Comment  
l'analyser ?

Entrée par  
les affects ?

# Témoignages d'étudiants

Plusieurs n'ont pas eu  
l'impression de  
travailler la langue  
mais de découvrir un  
outil

Côté affects : F. Baider,  
S. Coffey...  
Côté représentations :  
M. Dabène, J. Billiez...

Serions-nous dans la  
« bienveillance » (les gestes, les  
dispositifs) ?

Appel à contribution : *Mieux  
vivre en langues, ou comment  
passer de l'insécurité à la  
bienveillance, la bienveillance,  
la coopération*

# Réflexions à développer

## Côté étudiants

### Réflexion autour des **objectifs des étudiants**

- Rédiger leurs écrits académiques
- Devenir autonome dans la recherche

## Côté enseignants

### Réflexion autour des **outils nécessaires**

- Corpus en ligne : des exemples
- Langue spécifique : lexique transdisciplinaire

### Réflexion autour du **dispositif : bientraitance**

- Entrer par le « déjà là » : recherche en ligne
- Conduire vers des outils d'aide à la rédaction
- Conduire vers la langue spécialisée



# شكرا - Merci

Aide à l'écriture scientifique : une entrée outillée

**Cristelle CAVALLA**

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

