

Analyse du discours

Licence 2 : Lettres modernes

Séance 1 (18/01/23) : introduction

marine.delaborde@cyu.fr



CERGY PARIS
UNIVERSITÉ

Présentations

- <https://www.cyu.fr/marine-delaborde>
- À vous : <https://forms.gle/yKTBKuWMvQLRCWfn6>
 - À remplir pour la semaine prochaine

Organisation des séances

1 / 6

Séances	Points abordés
Séance 1 : 18/01/23 (salle 207)	Introduction
Séance 2 : 22/02/23 (salle 121)	Le projet, iTrameur et le corpus
Séance 3 : 22/03/23 (à confirmer) (salle 121)	Ventilation et concordance
Séance 4 : 29/03/23 (à confirmer) (salle 121)	Spécificités
Séance 5 : 5/04/23 (à confirmer) (salle 121)	Cooccurrents et segments répétés
Séance 6 : 12/04/23 (à confirmer) (salle 121)	Retour sur le projet

Objectifs du cours

- Voir syllabus

Introduction

Terminologie et concepts généraux

- Quel est le rapport entre les éléments suivants ?
 - Analyse du discours
 - Linguistique de corpus
 - Textométrie

Qualitatif / quantitatif

Précisions

- Articulation de données **qualitatives** et **quantitatives** :
 - **Analyse qualitative** :
 - Analyse fine, en profondeur (retour au texte)
 - **Analyse quantitative** :
 - Prise en compte d'un grand nombre de données à l'aide d'outils informatiques (mesures statistiques, tri)

Analyse du discours

Une approche multidisciplinaire

- **Discipline des sciences humaines et sociales (SHS) qui étudie le discours (contenu + contexte)**
- Méthodologie et concepts issus de différentes disciplines : sociologie, psychologie, histoire, sciences de la communication, linguistique, statistiques textuelles, informatique.
- Étude des marques linguistiques mais aussi de la situation d'énonciation
- **Approche qualitative et quantitative**
 - **Investigations sur corpus**

Linguistique de corpus

L'étude des données attestées

- Branche de la linguistique qui étudie le langage à travers des exemples réels
- Objectif de description et de modélisation de la langue
- Linguiste de corpus vs linguiste de fauteuil
 - Vieu débat sur la pertinence de travailler avec des corpus
- Perspectives de traitement automatique du langage naturel (TALN)
 - Création de corpus d'apprentissage pour réaliser des outils de traduction, extraction d'informations, chatbot (ex : ChatGPT), synthèse vocale, etc.
- Approche qualitative et quantitative
 - Investigations sur corpus

Textométrie / Statistiques textuelles (Pincemin, 2011)

Analyse statistique de données textuelles

- **Approche centrée sur le texte**
 - Mobilisée par différentes sciences humaines (histoire, littérature, etc.)
 - Pas incompatible avec la linguistique de corpus et l'analyse du discours !
- Caractérisée par certains **calculs fondateurs**, implémentés dans différents logiciels informatiques
 - Mesures statistiques : spécificités, cooccurrences.
 - Mesures non statistiques : segments répétés, concordances.
- Place importance donnée au « **retour au texte** »
 - pour l'interprétation des unités sélectionnées (mots) par les calculs
- **Approche quantitative et qualitative**
 - **Investigations sur corpus**

Textométrie / Statistiques textuelles (Pincemin, 2011)

Analyse statistique de données textuelles

- **Approche centrée sur le texte**
 - Mobilisée par différentes sciences humaines (histoire, littérature, etc.)
 - Pas incompatible avec la linguistique de corpus et l'analyse du discours !
- Caractérisée par certains **calculs fondateurs**, implémentés dans différents logiciels informatiques
 - Mesures statistiques : spécificités, cooccurrences.
 - Mesures non statistiques : segments répétés, concordances.
- Place importance donnée au « **retour au texte** »
 - pour l'interprétation des unités sélectionnées (mots) par les calculs
- **Approche quantitative et qualitative**
 - **Investigations sur corpus**

Les séries temporelles (Salem, 1991)






























Application textométrique

- De travailleurs à salariés. Repères pour une étude de l'évolution du vocabulaire syndical (1971-1990) (Salem, 1993)
 - Corpus :
 - Dans l'article :
 - Textes de 6 congrès (1973-1988) de la CFDT
 - 1 partition chronologique (en années)
 - Disponible sur iTraqueur [CFDT](#) :
 - Plus de textes (1973-2006)
 - Annoté automatiquement avec UDpipe (lemmes, POS, syntaxe)
 - 1 partition chronologique (en années)
 - Niveaux d'analyse des mesures textométriques :
 - Linguistique
 - Sociolinguistique
 - Juridique
 - ...

Les séries temporelles (Salem, 1991)

Application textométrique

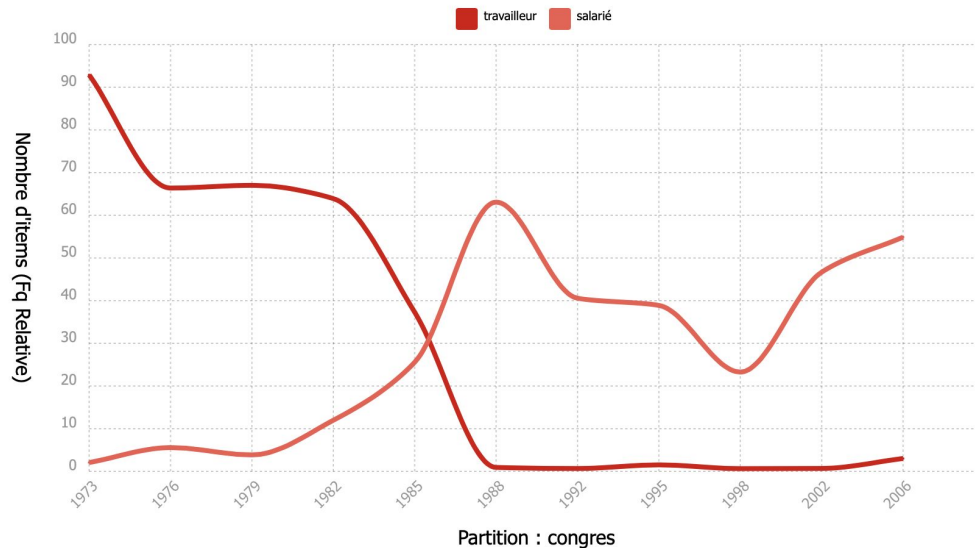
- De travailleurs à salariés. Repères pour une étude de l'évolution du vocabulaire syndical (1971-1990) (Salem, 1993)
 - Corpus : [CFDI](#)
 - Segments répétés :

SR	Fq	Lg	Concordance	Ventilation	Carte
à la	584	2			
à l'	433	2			
d' une	419	2			
les travailleurs	417	2			
et de les	395	3			
, les	390	2			
et les	389	2			
dans les	383	2			
les salariés	372	2			
le travail	368	2			

Les séries temporelles (Salem, 1991)

Application textométrique

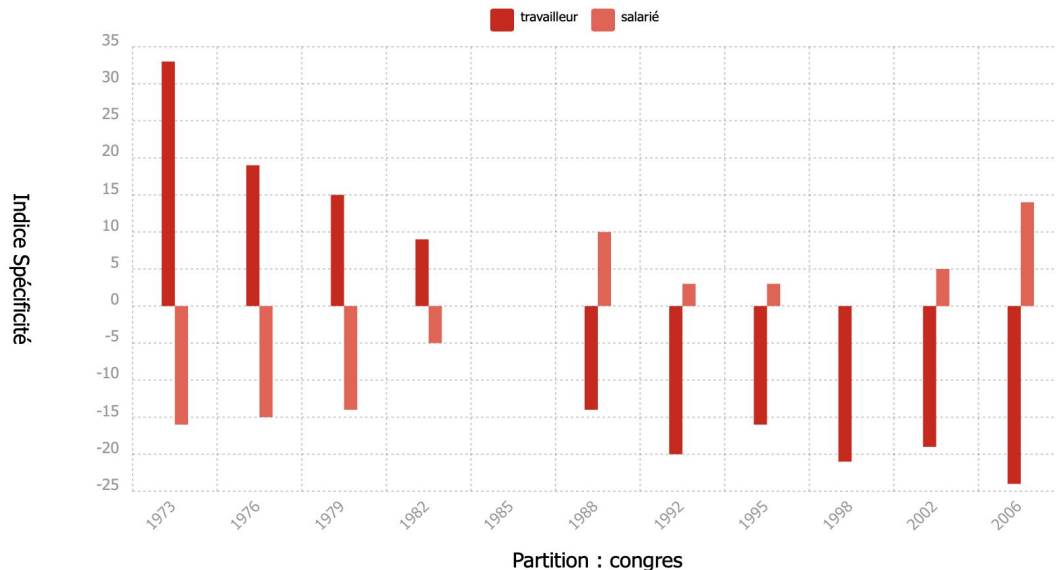
- De travailleurs à salariés. Repères pour une étude de l'évolution du vocabulaire syndical (1971-1990) (Salem, 1993)
 - Corpus : [CFDI](#)
 - Fréquences relatives (lemmes) :



Les séries temporelles (Salem, 1991)

Application textométrique

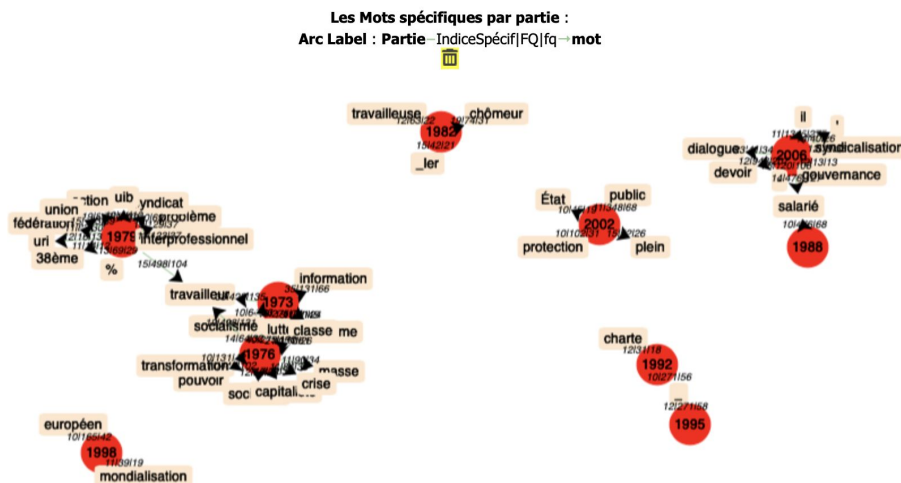
- De travailleurs à salariés. Repères pour une étude de l'évolution du vocabulaire syndical (1971-1990) (Salem, 1993)
 - Corpus : [CFDT](#)
 - **Spécificités (lemmes) :**
 - **Emplois substantifs**
(vérification à l'aide d'une concordance : précédés par des déterminants)



Les séries temporelles (Salem, 1991)

Application textométrique

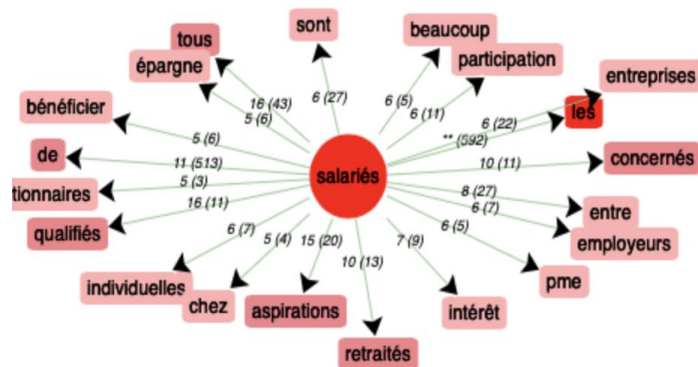
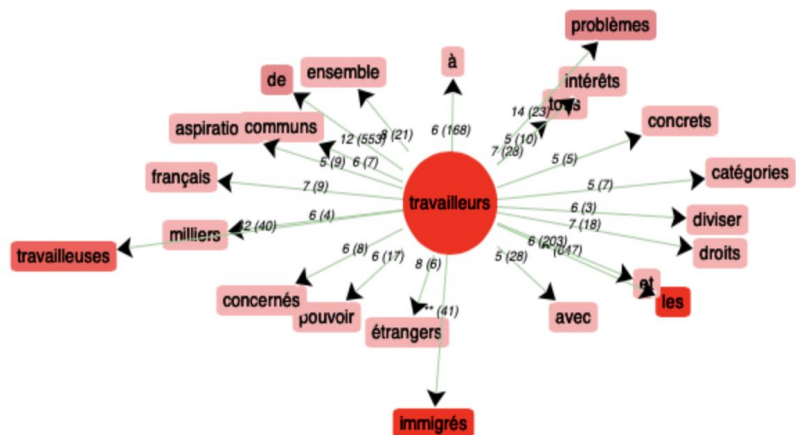
- De travailleurs à salariés. Repères pour une étude de l'évolution du vocabulaire syndical (1971-1990) (Salem, 1993)
 - Corpus : [CFDI](#)
 - Spécificités et cooccurents :



Les séries temporelles (Salem, 1991)

Application textométrique

- De travailleurs à salariés. Repères pour une étude de l'évolution du vocabulaire syndical (1971-1990) (Salem, 1993)
 - Corpus : [CFDI](#)
 - Cooccurents :



Les séries temporelles (Salem, 1991)

Application textométrique

- De travailleurs à salariés. Repères pour une étude de l'évolution du vocabulaire syndical (1971-1990) (Salem, 1993)
 - Corpus : [CFDI](#)
 - **Dépendance syntaxique :**
 - Relation NMOD
lutte / travailleur (lemmes)
 - Retour au texte :
concordance

Afficher 10 items Recherche :

N°	Partie	Contexte Gauche	Pôle	Contexte Droit
1	1973	la lutte anticapitaliste internationale elle recherche les possibilités de développer les	luttes	communes avec les travailleurs d' europe . § elle met tout
2	1973	commerce international ; § affirme toutefois qu' il considère comme légitime la	lutte	de les travailleurs de le tiers-monde contre tout régime d' oppression
3	1976	pratique d' élaboration de les revendications et de les formes de	lutte	avec les travailleurs et les travailleuses sont de les éléments essentiels
4	1976	indépendance active de la cfdt . § pour la cfdt , la	lutte	de les travailleurs a une dimension politique , qu' elle assume
5	1976	en certaines circonstances , son rôle est déterminant pour globaliser la	lutte	de les travailleurs face à l' adversaire de classe . § les
6	1976	pour une société socialiste autogérée . § la cfdt fait siennes les	luttes	de les travailleurs , cadres , enseignants et chercheurs , qui
7	1976	syndicales , problèmes et actions spécifiques d' un secteur dans la	lutte	globale de les travailleurs et les perspectives de la cfdt ;
8	1979	notre faveur dans notre pays . § elle dépend aussi de la	lutte	de les travailleurs de les autres pays industrialisés et de les
9	1979	Confédération ; § _assurer le soutien de l' opinion publique à les	luttes	de les travailleurs , et développer la propagande de masse par
10	1979	à le niveau européen et mondial . § cette action recouvre les	luttes	de les travailleurs dans les sociétés multinationales mais aussi celles menées

Affichage de 1 à 10 des 11 items Préc. 1 2 Suiv.

Les séries temporelles (Salem, 1991)

Application textométrique

- De travailleurs à salariés. Repères pour une étude de l'évolution du vocabulaire syndical (1971-1990) (Salem, 1993)
 - Corpus : [CFDT](#)
 - On trouve le plus souvent *immigré* dans le contexte droit de *travailleur*, ce n'est pas le cas pour *salarié*.
 - La substitution n'est pas systématique.
 - Distinction *travailleurs et travailleuses* vs *salarié et salariées* (« plan phonique »)
 - Connotations différentes associées à ces deux termes.
 - Statut différent dans le **domaine juridique** :
 - **Travailleur** : « n'a jamais constitué un concept juridique »
 - **Salarié** : « renvoie à une forme particulière d'emploi à laquelle sont rattachées un nombre considérable de dispositions juridiques de toute nature ». Le salarié est défini, en droit, par un contrat de travail.
 - Connotation liée au CDI associée au terme *salarié*
 - Salem retrouve des tendances similaires pour ces deux termes dans les corpus d'autres syndicats

Références bibliographiques

- Fillmore, C. J. (1992). Corpus linguistics or computer-aided armchair linguistics. In *Directions in corpus linguistics. Proceedings of Nobel Symposium* (Vol. 82, pp. 35-60).
- Pincemin, B. (2011). Sémantique interprétative et textométrie–Version abrégée. *Corpus*, (10), 259-269.
- Salem, A. (1993). De travailleurs à salariés. Repères pour une étude de l'évolution du vocabulaire syndical (1971-1990). *Mots. Les langages du politique*, 36(1), 74-83.
- Wallis, S., & Nelson, G. (2001). Knowledge discovery in grammatically analysed corpora. *Data Mining and Knowledge Discovery*, 5(4), 305-335.