جامعة البشير الإبراهيمي
UNIVERSITE EL BACHIR IBRAHIMI
BORDJ BOU ARRERIDJ

UNIVERSITE EL BACHIR IBRAHIMI. BORDJ BOU ARRERIDJ.

FACULTE DES LETTRES ET LANGUES.

DEPARTEMENT DE FRANÇAIS.

MASTER 1. Spécialité : LINGUISTIQUE

MODULE : LINGUISTIQUE DE CORPUS

Assuré par : Dre MILOUDI Imene, MCA , SCIENCES DU LANGAGE.

ANNEE UNIVERSITAIRE : 2025/2026

FICHE DESCRIPTIVE

PRESENTATION DU COURS

Ce cours est conçu pour les étudiants de master, futurs linguistes, et futurs enseignants. Nous y abordons les concepts-clés de linguistique de corpus comme la représentativité, l'échantillonnage, ainsi les types de corpus linguistiques, les principaux enjeux liés à la constitution, l'exploitation de corpus dans des travaux de recherche.

Nous plongerons dans l'analyse de données textuelles à partir de corpus authentiques (Corpus urbain, corpus de presse, corpus littéraire, corpus didactique, corpus sociolinguistique...etc) pour doter les étudiants des clés de la recherche en linguistique de terrain.

OBJECTIFS DU COURS ET COMPETENCES VISESEES

- La linguistique de corpus offre une approche empirique de l'étude des langues, en s'appuyant sur des données textuelles.
- Ce cours permettra de découvrir ce que c'est la linguistique de corpus et ses différentes utilisations dans l'analyse de discours, la linguistique appliquée. Il permettra de maîtriser les concepts fondamentaux et les méthodes d'analyse de corpus .
- L'étudiant de master apprendra à constituer un corpus de données langagières orales ou écrites en vue de l'analyse d'un problème ou phénomène linguistique (lexical, phonétique, sémantique, sociolinguistique, discursif...etc)
- L'étudiant de « master linguistique » aura l'occasion également de manipuler différents outils d'exploration de corpus afin d'entreprendre un travail de recherche empirique.
- S'exercer à mener une étude quantitative et une étude qualitative sur divers corpus.(urbain, journalistique, médiatique, didactique, littéraire...etc)

OBJECTIFS SPECIFIQUES au TD

1. Concevoir un corpus en vue d'une question de recherche spécifique.
2. Recueillir les données (orales ou écrites) + ou métadonnées
3. Transcrire les données ou les textes d'un corpus (transcription, tri, etc.)
4. Poser les questions de recherches / formuler des hypothèses ;
5. Apprendre à exploiter un corpus selon une méthode précise
6. Traitement et analyse de corpus authentiques en séances de TD.

PREREQUIS OU CONNAISSANCES PREALABLES

- IMPORTANT : Maitrise suffisante de la langue française.
- Notions de base de la linguistique (théories linguistiques, syntaxe, sémantique, lexicologie, sociolinguistique, pragmatique, analyse de discours)

METHODES D'ENSEIGNEMENT

Matière annuelle d'unité de méthodologie se présentant en :

- cours magistraux + Des travaux dirigés TD (4 crédits et coefficient 02)
- Séances consacrées à la lecture de travaux de recherches et des corpus variés en linguistique en Semestre1, puis la réalisation d'un travail de recherche personnel ou avant projet (ébauche de mémoire) en fin du Semestre2.
- Des séances de travaux dirigés consacrés à l'adaptation de mémoires académiques en mémoires « entreprise économique » ou startup pour préparer les étudiants souhaitant mener un projet entrepreneurial en 2ème année master.

MODALITES D'EVALUATION PONDERATION : 60% COURS + 40% TD

- Participation aux cours requise.
- Lecture et synthèse des corpus extraits de mémoires « linguistique » ou d'articles (Extrait d'ASJP).
- Remise d'un court travail (enregistrement, annotation, constat sur un phénomène linguistique) suite aux séances de lectures des travaux de recherche en TD .
- Travail écrit de 5 pages, avec bibliographie, mais sans les annexes à remettre en fin du semestre 2 .

NB. L'évaluation continue sera toujours prise en compte tout au long des deux semestres.

SUPPORT DU COURS

L'ensemble des synthèses de cours disponibles sur Moodle + Liens vidéo disponibles sur Moodle.

ENSEIGNANT INTERVENANT



Dre. MILOUDI Imene est maitresse de conférences au département de Français à l'Université de Mohamed El Bachir Ibrahimi. Elle a occupé le poste de responsable de filière durant la période (2015-2017) et la fonction de présidente du comité scientifique du département durant la période (2019-2022). Responsable scientifique et organisatrice de plusieurs journées d'étude et du colloque national "Pratiques et enseignement de l'Oral à l'Université algérienne : Bilan et perspectives." tenu le 10/5/2022.

Elle a notamment œuvré à l'ouverture de master « linguistique ». Ses recherches portent essentiellement sur les phénomènes de contact des langues en contexte plurilingue, l'étude des représentations en rapport avec le discours (épilinguistique, littéraire, médiatique, numérique...etc). Elle propose plusieurs enseignements autour de la linguistique générale, la sociolinguistique, l'analyse de discours et de la linguistique outillée en recourant aux méthodes de la linguistique de corpus, en particulier à l'aide de corpus analysés empiriquement.

TABLES DES MATIERES

PREAMBULE 1
CHAPITRE I : QUELQUES PRESUPPOSES THEORIQUES 4
Objectifs du chapitre I 5
COURS 01 : Les postulats de base d'une recherche empirique (scientifique) 6
1. DEFINITION D'UNE RECHERCHE SCIENTIFIQUE 7
2. LES POSTULATS DE BASE D'UNE RECHERCHE SCIENTIFIQUE 7
3 . LES TROIS NIVEAUX DE RECHERCHE 8
4 .OBJECTIFS DE LA RECHERCHE 9
5 QU'EST-CE QU'UNE RECHERCHE INTERESSANTE ? 9
COURS 02 : Etapes d'une recherche empirique à base de corpus : la phase exploratoire 11
1. LA QUESTION DE DEPART 11
2. LA PHASE EXPLORATOIRE 12
COURS 03 : Etapes d'une recherche empirique à base de corpus : Problématique et hypothèses de recherche 14
1.Définition de la problématique 14
2.Construction de la problématique 14
3.LA CONSTRUCTION DU MODELE D'ANALYSE (cadre d'analyse) 15
COURS 04 : Contexte épistémologique de la linguistique de corpus 18
1. UNE LINGUISTIQUE A BASE DE CORPUS 18
2. DEFINITION DE « CORPUS » 19
COURS 05 : Types et fonction des corpus. 22
1. Types de corpus 22
2. ROLE ET FONCTION DES CORPUS 24
CHAPITRE II : CONSTITUTION ET EXPLOITATION D'UN CORPUS 26
Objectifs du chapitre II 27
COURS 06: Constituer un corpus : Critères de recueil. 28
1. CRITERES DE SELECTION DE CORPUS (PROPRIETES RECHERCHEES) 29
COURS07 : Difficultés de constitution de corpus. 35
DIFFICULTES DE CONSTITUTIONS DES CORPUS. 35
COURS 08 : Principaux Outils de collecte des données : L'observation. 38
1. L'OBSERVATION OU (L'observation en situation) 38

2.Comment réaliser une observation ? 41
3. Avantages et limites de l'observation 41
4.Le recueil des données existantes : 43
COURS 09: Outils et techniques de collecte des données : Le questionnaire. 44
1. Le questionnaire : éléments de définition 44
2.Types de questionnaires 45
3.Critères d'élaboration d'un questionnaire 46
4.La structure d'un questionnaire 47
COURS 10 : Outils et techniques de collecte des données : L'entretien . 50
1. Définition 50
2. Types d'entretien 51
COURS 11 : Outils et techniques de collecte des données : L'analyse de contenu. 54
1.L'ANALYSE DE CONTENU 54
COURS 12 : OUTILS, TECHNIQUES DE COLLECTE DE DONNEES EN LINGUISTIQUE : L'expérimentation 57
1. Définition et fondements de la méthode expérimentale 57
2. Spécificité de la démarche expérimentale en linguistique 59
3. Les variables dans l'expérimentation 59
4. Le contrôle expérimental et l'analyse des résultats 59
5. Les étapes de la démarche expérimentale 60
Chapitre III: TRAITEMENT DES DONNEES DE CORPUS : MTHODES ET APPROCHES D'ANALYSE EN LINGUISTIQUE 61
Objectifs du chapitre III 62
COURS 13: Traitement des données : Structurer, annoter et analyser un corpus 63
1. De la collecte à la structuration du corpus 63
2. Annotation du corpus et enrichissement des données 64
3. Approches analytiques : entre description structurale et contextualisation 64
4. L'analyse du discours et l'intégration du contexte 64
5. Analyse qualitative des données de corpus 65
6.La synthèse interprétative 65
COURS 14: La Synthèse Interprétative En Linguistique De Corpus : Une Approche Systémique 67
1. Définition de la synthèse interprétative 67
2.Etapes et enjeux de la synthèse interprétative 68

3. Les deux niveaux de l'interprétation 69

COURS 15 : L'analyse Quantitative Des Corpus Linguistiques 72
1. Description des données linguistiques 72

2. Analyse des relations entre variables 72
3. Comparaison et interprétation des écarts 73

4. Limites de l'analyse quantitative 73

5. Vers une approche mixte (quantitative et qualitative) 73

COURS 16 : Panorama des approches linguistiques : l'approche communicationnelle (modèle de Jakobson) 75
1. L'approche communicationnelle : un modèle de Jakobson 75

2. Composantes et fonctions du langage 75

3. Exemples de l'analyse communicationnelle des corpus politiques 78

COURS 17 : L'approche énonciative : concepts de base 79
1. Définition et fondements de l'approche énonciative 79

2. L'appareil formel de l'énonciation 80

3. Énoncé, énonciation et situation d'énonciation 83
4. Les modes d'énonciation : énoncé ancré et énoncé coupé 83

COURS 18 : L'approche lexicométrique. 87
1. Définition et évolution de la lexicométrie 87

COURS 19 : L'approche dialogique et sémiologique 89
1. L'approche dialogique : le langage comme interaction 89

2. L'approche sémiologique : analyse des signes et des images 90

3. Analyse sémiologique de l'image 91

TRAVAUX DIRIGES 92
-Objectifs Spécifiques Des Travaux Dirigés (Td) 93
- Progression méthodologique des travaux dirigés en linguistique de corpus 94
TD 01 : ETAPES D'UNE RECHERCHE EMPIRIQUE, A BASE DE CORPUS : La problématique 95

- TD02 : Quelques Recherche Empirique A Base De Corpus 99
TD 03 : LES HYPOTHESES DE RECHERCHE 102

TD 04 : DEFINITION DU CORPUS EN « ANALYSE CONVERSATIONNELLE » 106
TD 05 : CRITERES DE RECUEIL DE CORPUS 109
TD 06 : L'OBSERVATION 111
TD 07 : L'OBSERVATION EN LINGUISTIQUE DE TERRAIN 116

- TD 08 : L'enquête Par Questionnaire 119
- TD 09 : Entretien Et Questionnaire : Approche Comparative. 121
- TD 10: Outils, Techniques De Collecte Des Données En Linguistique 124
- TD 11 : Analyse Et Interprétation Des Données De Corpus 129
- TD 12: Etude De Corpus Varies « Extraits De L'asjp » 133
- TD13: Outils De Collecte Et D'analyse De Données :L'analyse De Contenu 135
- TD 14: Telle Thématique, Tel Corpus : Études De Cas. 138
- TD 15 : Telle Thématique, Tel Corpus : Études De Cas. 140
- TD16 : Etude De Corpus Varies 143
- TD17 : du mémoire académique au mémoire strat up : des corpus a adapter 149
-CONCLUSION GENERALE 150
- REREFERENCES BIBLIOGRAPHIQUES 153

PREAMBULE

> *« Le corpus : quelle belle idée ! À condition que l'on veuille bien lire dans le corpus le corps : soit que dans l'ensemble des textes retenus pour l'étude (et qui forme le corpus), on recherche, nonplus seulement la structure, mais les figures de l'énonciation ; soit qu'on ait avec cet ensemble quelque rapport amoureux (faute de quoi le corpus n'est qu'un imaginaire scientifique).*
>
> **R. BARTHES, *Roland Barthes*, p. 163.**

La linguistique contemporaine connaît, depuis plusieurs décennies, un tournant empirique majeur qui redéfinit en profondeur les modalités d'analyse des faits de langue. Longtemps dominée par des approches introspectives ou théoriques, la discipline s'est progressivement orientée vers l'exploitation de données attestées, inscrites dans des contextes réels d'usage. C'est dans ce cadre que s'impose la linguistique de corpus, envisagée non seulement comme un ensemble de techniques, mais comme une véritable posture épistémologique fondée sur l'observation, la description et l'interprétation de données linguistiques authentiques. À ce titre, le corpus n'est plus un simple support illustratif : il devient un objet scientifique central, construit, interrogé et analysé selon des protocoles rigoureux.

La réflexion inaugurale de Roland Barthes, placée en exergue de ce module, souligne avec acuité la dimension presque « vivante » du corpus, envisagé comme un « corps » traversé par des dynamiques énonciatives. Cette métaphore invite à dépasser une conception strictement quantitative ou structurale du corpus pour en explorer les dimensions discursives, pragmatiques et interprétatives. En effet, travailler sur corpus implique de reconnaître que les données linguistiques ne sont jamais neutres : elles sont situées, contextualisées, porteuses de significations sociales, culturelles et idéologiques. Ainsi, l'analyse de corpus engage le chercheur dans une relation étroite avec ses données, nécessitant à la fois rigueur méthodologique et sensibilité interprétative.

Le présent module de « Linguistique de corpus » s'inscrit dans cette perspective, en proposant aux étudiants de master une formation approfondie aux fondements théoriques et aux pratiques méthodologiques de l'analyse empirique. Il vise à doter les futurs linguistes et enseignants des compétences nécessaires pour concevoir, constituer et exploiter des corpus dans le cadre de recherches scientifiques. L'accent est mis sur la compréhension des

principes fondamentaux tels que la représentativité, l'échantillonnage, la pertinence des données ou encore la structuration et l'annotation des corpus. Ces notions constituent des préalables indispensables à toute démarche de recherche fondée sur des données.

Au-delà de ces fondements, ce module ambitionne également de familiariser les étudiants avec la diversité des corpus linguistiques. Qu'il s'agisse de corpus oraux ou écrits, de corpus spécialisés ou généralistes, de données issues de la presse, de la littérature, des interactions quotidiennes ou des environnements numériques, chaque type de corpus répond à des objectifs spécifiques et soulève des enjeux particuliers. L'étude de ces différents formats permet de saisir la richesse et la complexité des pratiques langagières, tout en développant une capacité d'adaptation méthodologique en fonction des problématiques de recherche.

L'organisation du cours reflète cette double exigence théorique et pratique. Le premier chapitre, consacré aux présupposés théoriques, introduit les étudiants aux fondements de la recherche empirique. Il aborde notamment les postulats scientifiques, les étapes de construction d'une problématique, ainsi que le cadre épistémologique dans lequel s'inscrit la linguistique de corpus. Cette entrée en matière est essentielle pour comprendre que toute analyse de corpus repose sur une réflexion préalable concernant les objectifs de recherche, les hypothèses formulées et les méthodes mobilisées.

Le deuxième chapitre porte sur la constitution et l'exploitation des corpus. Il met en lumière les différentes étapes du recueil de données, en insistant sur les critères de sélection, les difficultés rencontrées sur le terrain et les outils disponibles pour la collecte d'informations. L'observation, le questionnaire, l'entretien ou encore l'analyse de contenu y sont présentés comme autant de techniques complémentaires permettant d'accéder à des données variées. Cette partie du cours accorde une attention particulière aux enjeux méthodologiques liés à la validité et à la fiabilité des données, conditions indispensables à toute recherche scientifique rigoureuse.

Le troisième chapitre s'intéresse au traitement et à l'analyse des données de corpus. Il propose un panorama des méthodes et approches mobilisées en linguistique, allant de l'analyse quantitative à l'analyse qualitative. Les étudiants y découvrent les principes de structuration et d'annotation des corpus, ainsi que les outils d'analyse permettant d'explorer les régularités linguistiques. Par ailleurs, différentes approches théoriques sont mobilisées – énonciative, pragmatique, lexicométrique, dialogique ou sémiologique – afin de montrer comment un même corpus peut être interrogé selon des perspectives variées. Cette pluralité

d’approches contribue à enrichir l’interprétation des données et à affiner la compréhension des phénomènes étudiés.

Enfin, une place importante est accordée aux travaux dirigés, conçus comme un prolongement pratique des enseignements théoriques. À travers des exercices progressifs, les étudiants sont amenés à concevoir leur propre corpus, à collecter et analyser des données, et à formuler des hypothèses de recherche. L’axe « Telle thématique, tel corpus ! » illustre de manière concrète l’importance de l’adéquation entre la problématique et le choix des données. En effet, un corpus pertinent est celui qui répond de manière cohérente à une question de recherche clairement définie.

Ce module se veut ainsi à la fois formateur et professionnalisant. Il prépare les étudiants à la réalisation de leurs travaux de recherche, notamment dans le cadre de leur mémoire de master, en leur fournissant les outils conceptuels et méthodologiques nécessaires. Plus largement, il les initie à une démarche scientifique rigoureuse, fondée sur l’observation, l’analyse et l’interprétation de données empiriques. Dans un contexte où les données linguistiques sont de plus en plus accessibles, notamment grâce aux technologies numériques, la maîtrise des méthodes de la linguistique de corpus apparaît comme une compétence essentielle pour les chercheurs en sciences du langage.

CHAPITRE I : QUELQUES PRESUPPOSES THEORIQUES

Objectifs du chapitre

Ce premier chapitre vise à installer les bases épistémologiques et méthodologiques de la recherche en linguistique de corpus. Il s'agit de permettre aux étudiants de comprendre les fondements scientifiques qui sous-tendent toute démarche empirique.

- Comprendre les postulats fondamentaux d'une recherche scientifique en sciences du langage.
- Identifier les étapes essentielles d'une recherche empirique fondée sur des données de corpus.
- Distinguer les notions clés : problématique, hypothèses, objet de recherche et cadre théorique.
- Situer la linguistique de corpus dans son contexte épistémologique et disciplinaire.
- Connaître les différents types de corpus et leurs fonctions dans la recherche linguistique.
- Développer une posture réflexive sur la construction de l'objet scientifique en linguistique.

COURS 01 : LES POSTULATS DE BASE D'UNE RECHERCHE EMPIRIQUE(SCIENTIFIQUE)

INTRODUCTION

À travers la recherche empirique, l'étudiant tend à explorer sa spécialité « la linguistique », en s'appuyant sur certaines méthodes auxquelles les étudiants master n'ont pas été directement confrontés en licence.

Toute **recherche universitaire** doit être empirique et s'articule autour de trois axes principaux : une **théorie, un corpus et une pratique.** Nous considérons que ces trois éléments sont ses piliers et même la condition sine qua non à son élaboration. Comment s'articule et se définit le rapport entre ces trois éléments, plus particulièrement entre la théorie et la pratique de manière à permettre à sélectionner un corpus donné au lieu d'un autre ?

À travers ce chapitre dédié à la notion et aux étapes d'une recherche scientifique, nous visons à initier l'étudiant la recherche empirique, faire en sorte qu'il soit en mesure de formuler une question de recherche et de déterminer des objectifs et des hypothèses ; d'élaborer et conduire de manière rigoureuse une activité de recherche sur un phénomène linguistique particulier en testant et administrant des outils de collecte de données quantitatives et qualitatives.

Mener un travail de recherche sur un sujet précis en se servant d'un corpus commence à partir *de « l'idée de recherche » jusqu'au « traitement et l'analyse des données collectées : le corpus » : autrement dit* , conduire l'étudiant depuis *la définition du problème (limites et exigences) à étudier jusqu'à l'élaboration de sa problématique d'enquête,* la collecte des informations brutes qu'on appelle corpus et la formulation de ses hypothèses de recherche.

2. DEFINITION D'UNE RECHERCHE SCIENTIFIQUE

Toute collecte de données, d'informations pour l'avancement du savoir peuvent être des activités de recherche, telles que l'explique d'ailleurs M. Le grain

> **«** *La recherche est un effort pour trouver quelque chose ou un effort de l'esprit vers la connaissance* » (1994 :945).

D. Bruno définit également la recherche comme :

> « *un exercice systématique et méthodique portant sur l'étude d'un problème ou d'une question et mettant en cause des faits qui doivent être vérifiables en vue d'atteindre une fin : la résolution d'un problème ou la réponse à une question ou d'une hypothèse préalable, la recherche exige ipso facto un travail d'interprétation* » (1994 :85) .

D'après BrunoD, la recherche ouvre aussi une porte sur l'inconnu et amène des solutions à de nombreux problèmes dans le monde.

De nombreux auteurs soulignent qu'il existe une panoplie de définitions de la recherche scientifique. Elles varient selon les domaines de recherche et la façon avec laquelle le chercheur conçoit la connaissance.

La recherche scientifique est définie comme un processus dynamique ou une démarche systématique d'acquisition de connaissances permettant ainsi d'examiner des phénomènes, et d'obtenir des réponses précises en s'appuyant sur **l'investigation**. La caractéristique principale de ce processus est sa systématicité et sa rigueur qui visent à répondre à une question ou à résoudre un problème, comme l'illustre :

Une recherche scientifique a plusieurs visées : décrire un phénomène, l'expliquer, contrôler et prédire des faits, des phénomènes et des conduites.

3. LES POSTULATS DE BASE D'UNE RECHERCHE SCIENTIFIQUE

3.1. ***Le déterminisme*** **:** la recherche vise à repérer les constantes qui déterminent la régularité des phénomènes, elle s'appuie sur le principe de causalité.

3.2. ***L'empirisme*** **:** la recherche s'appuie sur le recueil des faits qui organisent un phénomène, ce qui permettra établir des modèles.

3.3. ***L'intégration théorique*** **:** elle vise à déterminer les lois qui gèrent le fonctionnement des choses.

3.4. ***La dimension dynamique*** **:** elle s'appuie sur le principe de la perpétuelle

évolution. Il n'y a jamais de résultats stables et définitifs, mais des perspectives.

3.5. ***La dimension publique*** : la science vise à généraliser les résultats et les rendre publics.

4. LES TROIS NIVEAUX DE RECHERCHE

La recherche s'appuie essentiellement sur trois niveaux constitutifs :

4.1. La description

La description se donne pour objectif de caractériser la nature des phénomènes à étudier et de déterminer leurs points de partage. Elle peut être le centre d'une recherche : à titre d'exemple déterminer tous les aspects d'un processus ou d'une méthode d'apprentissage. Pour une recherche, la description peut occuper la première phase de l'analyse dans la mesure où elle se charge de présenter les résultats d'une observation ou d'une enquête exploratoire. Á ce niveau de la recherche, la présence d'une méthode rigoureuse et d'hypothèses bien élaborées, s'avère crucial.

*4.2. La **classification***

La classification permet de catégoriser, répartir, ordonner, afin de pouvoir établir des comparaisons ou des rapprochements. La matière analysée ou étudiée se trouve ordonnée, structurée, répartie en catégories et sous -catégories pour faciliter sa compréhension.

4.3. L'explication / compréhension

Comme son nom l'indique, ce niveau se présente comme une phase qui va apporter des réponses pertinentes à la question « *pourquoi ?* ». Autrement dit, il fournir des informations expliquant la genèse d'un phénomène et les causes qui ont fait qu'il soit arrivé à l'état actuel. Les explications doivent être en mesure de véhiculer des réponses clarifiant ainsi les relations entre des phénomènes et déterminant les conditions dans lesquelles certains phénomènes ou certains évènements ont pu avoir lieu.

5. OBJECTIFS DE LA RECHERCHE

La fonction principale de la recherche scientifique est de permettre au chercheur de mieux comprendre, de mieux expliquer et de mieux interpréter les phénomènes observés. En fait, c'est par le bais de ce processus scientifique qu'il sera en mesure :

- d'augmenter, d'enrichir son Savoir,
- de fournir une description aussi objective et complète que possible sur le phénomène abordé.
- de procéder à une prise d'informations, à un recueil de données à la fois pertinentes, objectives et aussi complètes que possible.

6. QU'EST-CE QU'UNE RECHERCHE INTERESSANTE ?

La définition scientifique de la recherche stipule qu'elle ne doit pas seulement être de qualité, elle devrait aussi être intéressante. En d'autres termes, elle doit contribuer à la réalisation d'une étude ***nouvelle.*** Plus précisément, le chercheur doit être en mesure :

- de produire quelque chose qui est *nouveau* d'une manière ou d'une autre.
- d'apporter des réponses à de nouvelles questions ou à d'anciennes questions sans réponses appropriées,
- de répondre avec plus de pertinence à des questions traitées dans la littérature,
- de soutenir des réponses existant déjà dans la littérature par une nouvelle argumentation.

A. Boaz et D. Ashby suggèrent que les «*conceptualisations de la qualité en recherche doivent être poussées au-delà de la seule qualité méthodologique et traiter du "caractère pertinent" de la recherche*» (2003 :12). Ces auteurs traitent des dimensions de la «pertinence», c'est-à-dire la «*pertinence des méthodes selon les objectifs de la recherche*» et «*la pertinence de la recherche en fonction de l'utilisation probable de ses constatations*».

En outre, à travers une recherche intéressante le chercheur peut apporter :

- de la satisfaction personnelle,
- quelque chose qui « fait plaisir » à une certaine communauté
- des réponses à des questions qui sont importante pour le développement d'un domaine scientifique,

- des réponses à des questions qui sont importantes pour l'amélioration d'un programme déjà établi,
- des réponses à des questions posées par la communauté scientifique.

Conclusion

Une recherche universitaire est donc, une investigation organisée rigoureuse ayant pour but un phénomène particulier. C'est pourquoi toute recherche est censée avoir un sujet précis, une problématique, un plan et une méthode. Tous ces éléments doivent être explicités par l'étudiant au début de son travail, car le processus de la recherche compte tout autant que les résultats auxquels il peut aboutir. Ce qui importe le plus, c'est l'adoption d'une démarche heuristique, rigoureuse et cohérente.

Quel que soit le type de sujet, la recherche ne doit pas être une simple compilation d'informations ni une synthèse de l'existant. Elle doit apporter du « nouveau », un surplus de connaissances, soit au niveau de la méthode, c'est-à-dire la manière de faire, soit enfin au niveau des résultats.

COURS 02 : LES ETAPES D'UNE RECHERCHE EMPIRIQUE A BASE DE CORPUS : la phase exploratoire.

INTRODUCTION

Quel que soit le sujet traité et en fonction du domaine, les étapes d'une recherche en linguistique tout comme en sciences humaines et sociales peuvent être résumées en plusieurs phases : Phase exploratoire, de documentation, phase d'analyse (l'étude) et phase de rédaction.

La recherche empirique se présente sous forme d'étapes interelliées, homogènes et cohérentes.

3. LA QUESTION DE DEPART

La question de départ est une interrogation précise qui constitue un point de départ pour le chercheur, et de retour (après un certain périple) à ce même point. Elle aide à délimiter l'axe de la recherche.

Elle se donne pour objectif de délimiter l'étendue de la recherche pour éviter au chercheur le gaspillage de temps et d'énergie, comme l'explicitent d'ailleurs Raymond Quivy et Luc Van Campenhoudt dans ce qui suit « *[...] énoncer son projet de recherche sous la forme d'une question de départ « par laquelle le chercheur tente d'exprimer le plus exactement possible ce qu'il cherche à savoir, à élucider, à mieux comprendre (1995 :12).*

De ce fait, pour constituer une question utile à son étude, le chercheur doit respecter trois critères lors de son élaboration : **la clarté** (précise, concise, compréhensible), la **faisabilité** (réalisable) et **la pertinence** (n'est pas métaphysique ni moralisatrice).

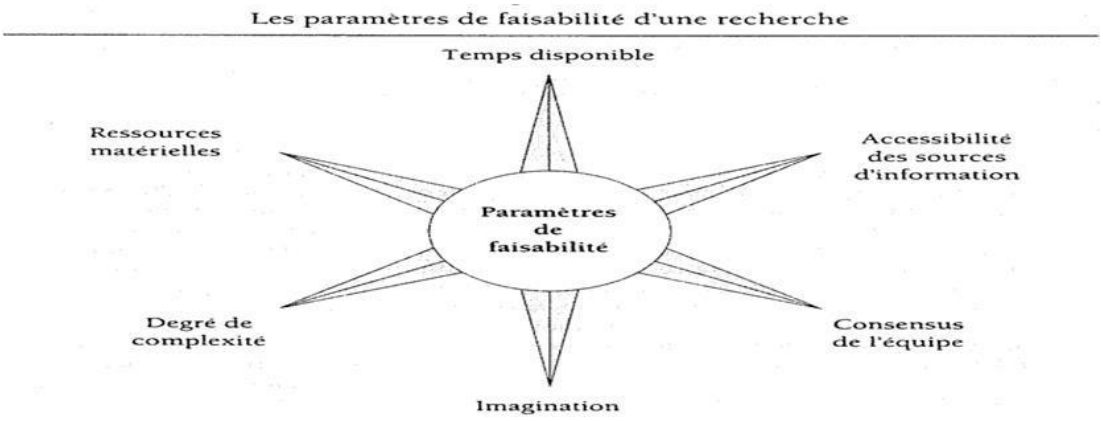


Figure 01 : Paramètre de la caractéristique : faisabilité (SMAIL Idir, 2018 : 10)

4. LA PHASE EXPLORATOIRE

Elle est considérée comme une phase préalable à la recherche, au cours de laquelle le chercheur met à jour un minimum de savoirs, clarifie un problème plus ou moins défini initialement et détermine de nouvelles pistes de recherche. Cette étape consiste donc à la recherche de données et d'informations, sur un terrain d'étude donné, afin de pouvoir circonscrire un objet de recherche et en conséquence choisir des théories ou opter pour une méthodologie propice à l'objet de recherche :

> « *L'exploration permet ainsi une émergence ou une mobilisation de théories qui vont conduire à la formulation de questions de recherche, de porter un regard différent, qui pourrait conduire à de nouvelles découvertes; de même, concernant un contexte déjà largement étudié qui aurait évolué*. » (V.Voisin, 2017 :30).

La lecture (revue de la littérature ou état de l'art)

Cette phase de lecture préparatoire est généralement placée sous la rubrique « *État de l'art* » et sert à se renseigner sur les travaux spécifiques déjà réalisés sur le sujet que le chercheur veut traiter. Pour cette phase, le chercheur va lire les ouvrages et les articles sur la question. Ainsi, le bon déroulement de cette phase implique le respect de certains critères :

a. Déterminer le lien des ouvrages ou articles choisis avec la question de départ

b. Indiquer avec précision la nature de la recherche, l'auteur, la date, l'université

ou l'école, le lieu, les hypothèses de départ.

c. Identifier les éléments d'analyse, la méthodologie suivie, les résultats obtenus et leur interprétation
d. Déterminer l'apport et les retombées de chaque recherche (prévues et réelles).

Cette étape est très importante dans la mesure où *elle permet à l'étudiant de s'inspirer d'approches et de méthodologies diverses appliquées à un même problème.*

Au terme de cette étape il aura une idée détaillée sur le plus essentiel de ce qui a été déjà fait, en théorie et en analyse, sur le même sujet.

Conclusion :

En définitive, la phase exploratoire apparaît comme un moment charnière de la recherche empirique, à la fois préparatoire et structurant. Elle permet de transformer une intention de recherche en un projet scientifique cohérent, en articulant de manière progressive la question de départ, la revue de la littérature et la construction de l'objet d'étude. Dans le cadre de la linguistique de corpus, cette étape revêt une importance particulière, car elle conditionne directement la qualité du corpus constitué, la pertinence des analyses effectuées et la validité des résultats obtenus. Cette étape favorise la mobilisation progressive de cadres théoriques et méthodologiques adaptés à l'objet d'étude. En confrontant différentes approches et en analysant la manière dont elles ont été appliquées à des problématiques similaires, le chercheur affine ses propres choix méthodologiques. Il développe ainsi une capacité de réflexion critique indispensable à toute démarche scientifique rigoureuse.

Ainsi, maîtriser la phase exploratoire revient à acquérir les bases essentielles d'une démarche scientifique rigoureuse, capable de produire un savoir fondé, argumenté et situé dans un champ disciplinaire clairement défini.

COURS 03 : ETAPES D'UNE RECHERCHE EMPIRIQUE A BASE DE CORPUS : PROBLEMATIQUE ET HYPOTHESES DE RECHERCHE

INTRODUCTION

Avant de procéder au choix de la technique d'enquête et à la formulation des hypothèses, le chercheur doit concevoir en amont un « *problème* » à élucider, à étudier lors de sa recherche. Cette étape est nécessaire pour le processus de recherche. Le chercheur élabore sa problématique après avoir « *déterminé ce qui fait problème* ».

1.Définition de la problématique

Selon R. Quivy et L-V. Campenhoudt, la problématique est « *l'approche théorique que l'on décide d'adopter pour traiter le problème posé par la question de départ* » (2017 : 105). Elle nécessite un effort de construction, d'agitation d'idées, de pensées, de théories sur ce qui fait problème dans un sujet.

À ce stade, tout chercheur doit s'interroger : « *En quoi a-t-on besoin d'effectuer cette recherche et vais-je apporter des éléments nouveaux à la science et à la pratique ?* ». En fait, cette étape exige de la créativité, de l'ingéniosité pour fournir des éléments justifiant la mise en œuvre de cette recherche.

2.Construction de la problématique

Une problématique de recherche est constituée d'un ensemble de conceptions théoriques, de lectures, de questions, d'hypothèses et de références qui aident à clarifier et à éclaircir un problème de recherche. À partir de l'ensemble de questions auquel le chercheur se trouve confronté, il convient de formuler une problématique en termes de questions pertinentes. Il s'agit de :

- déterminer le problème ou le phénomène observé.
- formuler un énoncé affirmatif qui explique la mise en place de cette recherche.
- élaborer une argumentation explicitant la nécessité et la pertinence de cette exploration empirique et ses retombées sur l'avancement scientifique.

Les auteurs proposent une méthode en trois temps pour la conception d'une

problématique. :

- **Dans un premier temps, faire le point :** cela signifie qu'il s'agit de dresser l'inventaire des problématiques éventuelles et de les comparer (déterminer les liens ou les oppositions) en fonction des résultats de l'exploration.
- **Dans un second temps, proposer une problématique ou un cadre théorique :** le chercheur doit être en mesure en ce moment de choisir son cadre théorique, soit en s'adaptant à celui proposé par des auteurs lus, soit en suggérant un nouveau. Ce qui importe dans tout les cas est qu'il soit en adéquation avec la question de départ et avec la revue de la littérature- lectures-
- **Dans un dernier temps**, expliciter sa propre problématique et son cadre théorique : au niveau de cette étape, le chercheur présente sa problématique et son cadre théorique après être reformulés. Dans ce cas le chercheur pourra construire son modèle d'analyse.

En résumé, trois étapes : formulation de la question de départ, exploration et enfin *explicitation de la problématique, constituent un système harmonieux pour la concrétisation d'un projet de recherche.*

3.LA CONSTRUCTION DU MODELE D'ANALYSE (cadre d'analyse)

Cette étape vient ***après la construction finale de la problématique pour permettre au chercheu***r de réaliser le travail élucidé. Elle en est donc indépendante. Cependant, il est plus logique de travailler à l'élaboration de cette étape en liaison et en complémentarité avec la précédente : elles doivent être presque concomitantes.

Le modèle d'analyse ou le cadre d'analyse se définit comme : **«** *un ensemble de concepts et d'hypothèses qui s'articulent entre eux pour former un cadre cohérent* **» (**R. Quivy et L-V. Campenhoudt , 2017 :153).

L'élaboration de ce cadre se réalise à travers l'élaboration d'hypothèses.

3.1.LES HYPOTHESES

L'hypothèse se définit comme une proposition admise provisoirement avant qu'elle soit confirmée, infirmée ou nuancée par la confrontation des faits. Pour

l'hypothèse est « *en quelque sorte une base avancée de ce que l'on cherche à prouver.*

C'est la formulation pro forma de conclusions que l'on compte tirer et que l'on va s'efforcer de justifier et de démontrer méthodiquement et systématiquement » (O. Aktouf 1987 :58). Elle remplit trois fonctions principales :

- Établir des liens entre des faits, des variables et des concepts pour parvenir à mettre en place une règle applicable à de nombreuses situations données.
- Orienter le choix des faits à observer, des données à collecter afin de pouvoir répondre aux questions centrales posées lors de la recherche.
- Fournir des renseignements précis quant au terrain de la recherche, ce qui permettra de choisir au final la méthode appropriée, les techniques pour confirmer ou infirmer les relations énoncées.

Reste à savoir ce qu'est une hypothèse correcte et acceptable dans la recherche scientifique.

3.2.Formulation des hypothèses

Les paramètres à prendre en considération en formulant des hypothèses :

- Il faut qu'elle soit liée à une théorie, une conception précédemment citée.
- Il faut qu'elle soit vérifiable : le principe d'une hypothèse est qu'elle soit en mesure d'être vérifiée en possédant des variables observables, mesurables dans la réalité et contestables.
    - S'il y en a plusieurs, il faut qu'elles restent en nombre limité.
    - La plausibilité : elle doit être pertinente pour le phénomène à étudier.

Les hypothèses constituent des points de départ pour votre recherche. Elles permettront aux étudiants chercheurs de construire un cadre théorique solide.

Pour affiner les hypothèses, on peut se poser les questions suivantes :

- Quels sont les concepts théoriques pertinents pour aborder cette problématique ?
- Quelles sont les études précédentes qui ont abordé des sujets similaires ?
- Quelles sont les données dont je dispose ou que je peux collecter pour tester mes hypothèses ?

Conclusion

Ce cours constitue une étape décisive dans le processus de la recherche empirique en

linguistique de corpus. En effet, il marque le passage d'une réflexion exploratoire générale à une structuration scientifique rigoureuse de l'objet d'étude, en posant les bases conceptuelles et méthodologiques nécessaires à toute investigation empirique.

La construction de la problématique apparaît ainsi comme un moment central de la démarche scientifique. Elle ne se limite pas à la formulation d'une simple question, mais engage un véritable travail de conceptualisation, d'analyse et de positionnement théorique. À travers l'exploration des lectures, la confrontation des idées et la mise en relation des approches existantes, le chercheur élabore un cadre interprétatif cohérent qui lui permet de donner sens à son objet d'étude. Ce processus progressif, allant de la question de départ à l'explicitation de la problématique, garantit la pertinence et la profondeur de la recherche.

Dans cette continuité, la construction du modèle d'analyse vient structurer davantage la démarche scientifique. Elle permet de traduire la problématique en un ensemble organisé de concepts et de relations théoriques, formant un cadre cohérent d'investigation. Ce modèle constitue ainsi une passerelle entre la réflexion théorique et l'analyse empirique, en orientant concrètement le travail du chercheur sur le terrain. Les hypothèses de recherche occupent, dans ce dispositif, une fonction essentielle. Elles représentent des propositions provisoires, formulées à partir de la problématique, et appelées à être confrontées aux données du corpus. Leur rôle est multiple : elles orientent la collecte des données, structurent l'analyse et permettent de tester la validité des relations supposées entre les phénomènes étudiés. À ce titre, elles constituent un véritable guide pour la recherche, en assurant la cohérence entre les objectifs théoriques et les résultats empiriques attendus.

Ainsi, la problématique et les hypothèses ne doivent pas être considérées comme de simples étapes formelles, mais comme le cœur même de la démarche scientifique. Elles traduisent la capacité du chercheur à problématiser le réel, à structurer sa pensée et à construire un savoir fondé sur l'observation et l'analyse rigoureuse des données linguistiques.

COURS 04 : CONTEXTE EPISTEMOLOGIQUE DE LA LINGUISTIQUE DE CORPUS

INTRODUCTION

Le corpus occupe un espace dans la recherche scientifique. A partir de la diversité des sa définitions, maintes questions surgissent sur l'essence du corpus et les mécanismes qui le régissent et qui le dominent. Les études menées par les linguistes francophones et anglo- saxons montrent que le corpus mérite d'être étudié sous différents angles.

Nous nous proposons donc à travers ce chapitre II, d'en étudier ses multiples aspects

3. UNE LINGUISTIQUE A BASE DE CORPUS

La branche de la linguistique qui se préoccupe plus spécifiquement des corpus s'appelle logiquement la linguistique de corpus. Elle est liée aujourd'hui au développement des systèmes informatiques, en particulier à la constitution de bases données textuelles.

Avec une recherche à base de corpus, l'étudiant interroge ses propres intuitions, pose ses propres hypothèses et les vérifie par l'analyse de données réelles. Les avantages d'une telle approche sont nombreux sur le plan de l'intérêt suscité chez l'étudiant qui est amené à faire des observations à partir de ses intuitions ; il devient alors acteur dans la construction de son propre savoir. Certes, le travail de l'étudiant ne pourra se faire qu'en semi-autonomie : il sera guidé par directeur de recherche dans la découverte de certains traits de la langue.

Mais quels sont les principes épistémologiques sur lesquels se basent cette discipline ?

Le courant britannique de la linguistique de corpus trouve ses fondements théoriques dans l'école contextualiste dont le précurseur est Firth. Son point de départ est le fait qu'aucun énoncé ne peut être compris en dehors du contexte culturel et situationnel dans lequel il est produit, sans tenir compte des informations d'ordre sociologique (âge, sexe, milieu) ainsi que des informations qui concernent l'acte énonciatif (genre de discours, endroit où il a été produit, buts de l'énonciation, etc.). D'où l'exigence de fonder ses analyses linguistiques sur les observations de « ce qui est visible » (Williams, 2003).

Les disciples de **Firth** ont contribué à diffuser ses idées innovantes, en particulier Sinclair qui a été le premier véritable maître à penser de la linguistique de corpus et qui a permis que cette discipline prenne réellement son essor. Son approche de la langue consiste à *l'observer, à faire des hypothèses sur son fonctionnement*.

Une autre proposition innovante de **Sinclair** consiste à considérer que ***les analyses linguistiques ne doivent pas se baser sur des textes particuliers, mais sur de grands ensembles de données***

dans la mesure où c'est seulement dans de très grands corpus que le chercheur peut trouver des régularités qui permettent de décrire le fonctionnement de la langue.

Certains spécialistes du domaine considèrent la linguistique de corpus comme une *méthodologie applicable à différentes disciplines* d'autres, au contraire, pensent que la linguistique de corpus *est une discipline à part entière dans la mesure où elle a des présupposés* ***épistémologiques évidents (***Sinclair, 1991 ; Leech, 1992 ; Tognini-Bonelli, 2001).

Cette dernière perspective d'analyse consiste, en effet, en une étude attentive et objective se basant sur des données linguistiques ; c'est donc une étude qui permet de généraliser les résultats obtenus à la totalité de la langue, à condition, bien sûr, que le type de corpus analysé le permette.

4. DEFINITION DE « CORPUS »

Aujourd'hui, tout chercheur conçoit le corpus comme un observable nécessaire en linguistique, l'épistémologie fondamentale de la discipline se joue et se rejoue, dans la réflexion sur les corpus . Traditionnellement, et même dans le vocabulaire des sciences, ***« corpus désigne un recueil large, et quelquefois exhaustif, de documents et de données ».***

Le concept a pourtant conservé, en linguistique, un peu de son acception d'origine. « *Corpus* » du latin *« corpore »* qui veut dire « corps » renvoie, en un premier sens, à une collection de textes présentant une certaine unité de genre ou bien d'époque.

Dans les sciences humaines et sociales tout particulièrement, *« corpus désigne les données servant de base à la description et l'analyse d'un phénomène » (*P. Charaudeau et al, 2002, P.148)

Un corpus offre au linguiste un certain nombre d'énoncés écrits et oraux pour lesquels le problème d'acceptabilité ne se pose pas. (Robert Galisson et al 1976, P.131)

Les énoncés oraux ont été réellement produits, sont donc attestés, même si les difficultés tenant à la transcription laissent planer un doute sur certains d'entre eux.

Le corpus est ensemble homogène et significatif de données linguistiques observées et à partir desquelles pourra s'élaborer la théorie.

Aussi, le corpus en (ling.) ensemble limité des éléments (énoncés) sur lesquels se base l'étude d'un phénomène linguistique ; en (lexicométrie), c'est un ensemble de textes réunis à des fins de comparaison, servant de base à une étude quantitative. (Lebart, Salem 1988, § *Glossaire*)

Pour le linguiste, définir un corpus, qui va être son objet d'étude, n'est pas chose facile. En effet, à la suite de Saussure, la linguistique s'est donné comme objet la langue. Il ne s'agit pas d'étudier

des usages particuliers, a priori infinis, mais bien de dégager ce qui fait système, les règles qui sous-tendent le fonctionnement langagier, d'une langue, voire de toutes les langues. Or, avec un corpus, on n'est plus dans un fonctionnement potentiel mais bien dans la réalité des usages, nécessairement corrélés à des situations de communication particulières.

Définir un corpus oblige à circonscrire un objet clos et donc à se donner des règles pour garantir un minimum de représentativité à cet objet, cette représentativité étant elle-même une condition de réussite de la recherche. Nous proposons ainsi plusieurs définitions au « corpus » dont les plus pertinentes :

J.Sinclair :

> *« Un corpus est une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage »*

(Sinclair cité par Habert et a, 1997 :144)

Rastier(2004) :

> *« Un corpus est un regroupement structuré de texte intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés(i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'application ».* *Rastier François, 2004 « Doxa et lexique en corpus-pour une sémantique des idéologies.*

Le corpus peut se définir comme un recueil de données des faits de langues. La linguistique de corpus a été développée pour extraire d'un corpus les connaissances linguistiques nécessaires à l'enseignement des langues ; un corpus étant un ensemble collecté et ordonné de données langagières réelles. La communauté linguistique considère, à la suite de Sinclair (1996), qu'un corpus est "*une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon de langage"* Habert, Salem 1997, p. 11)

Le recueil du corpus se fait en fonction des besoins de la recherche, selon les objectifs que tracent les chercheurs, selon (Rastier : 2005 : 45) *« si un corpus ne correspond pas à un besoin intellectuel ou scientifique, il se périme et perd toute sa valeur »*

CONCLUSION

Le contexte épistémologique de la linguistique de corpus met en évidence le passage d'une linguistique fondée sur l'intuition à une linguistique résolument empirique, centrée sur l'observation des usages réels de la langue. En s'inscrivant dans la tradition contextualiste et en

s'appuyant sur le développement des technologies informatiques, cette approche considère le corpus comme un objet scientifique à part entière, permettant de saisir le fonctionnement du langage dans sa dimension effective et située.

Ainsi, la linguistique de corpus se définit moins comme une simple méthode que comme une posture scientifique reposant sur l'analyse de données authentiques, structurées et représentatives. Elle offre au chercheur la possibilité de confronter ses hypothèses aux faits linguistiques observables, tout en favorisant la mise en évidence de régularités à grande échelle.

Par ailleurs, la réflexion sur la notion de corpus révèle sa complexité et sa centralité dans les sciences du langage. En tant qu'ensemble organisé de données langagières, le corpus constitue à la fois un objet construit et un outil d'analyse, dont la pertinence dépend étroitement des critères de sélection, de structuration et de représentativité.

En définitive, ce cours met en lumière le rôle fondamental du corpus dans la construction du savoir linguistique contemporain, en tant que support empirique indispensable à toute démarche scientifique rigoureuse et contextualisée.

COURS 05 : TYPES ET FONCTION DE CORPUS

INTRODUCTION

- Dans quels domaines utilise t-on les corpus?

Les corpus sont utilisés dans plusieurs domaines, parmi lesquels nos pouvons citer :

- La lexicographie (la constitution de dictionnaires).
- L'apprentissage des langues.
- Les études linguistiques et sociolinguistiques (l'étude de vocabulaire, de la grammaire, l'évolution de la langue ou des sens des mots, la linguistique informatique (Traitement Automatique des Langues Naturelles), ...).
- La terminologie, la traduction, la rédaction technique…

1. Types de corpus

Il existe plusieurs appellations pour désigner les corpus, selon les traditions disciplinaires et les objectifs de recherche. Ainsi, dans le domaine de la lexicographie, on parle souvent de « corpus sacs de mots », centrés principalement sur le lexique et les occurrences de mots. De leur côté, les grammairiens mobilisent des corpus constitués de phrases parfois hors contexte, mais jugées représentatives de la bonne formation des énoncés. Ces différents usages montrent que la notion de corpus varie en fonction des perspectives adoptées et des finalités analytiques.

Dans une perspective didactique, on distingue également les corpus d'apprenants, élaborés à partir des productions langagières d'individus en cours d'apprentissage d'une langue étrangère ou seconde. Ces corpus permettent non seulement d'observer les processus d'acquisition, mais aussi d'identifier les erreurs récurrentes afin d'améliorer les pratiques pédagogiques. Par ailleurs, on parle de corpus bruts pour désigner des ensembles de données non annotées, constitués de textes écrits ou de transcriptions de productions orales. D'autres dénominations existent encore, comme les corpus urbains, qui renvoient à des données recueillies en milieu urbain, ou encore les corpus littéraires et linguistiques, définis en fonction de leur contenu ou de leur visée analytique.

Les corpus peuvent être de nature orale ou écrite, bien que les corpus écrits soient historiquement les plus répandus en raison de leur facilité de collecte et de traitement. Toutefois, les évolutions technologiques ont favorisé le développement de corpus oraux, permettant une meilleure prise en compte des usages réels de la langue en interaction.

Enfin, si l'on considère le terme « corpus » dans son acception la plus large, comme un ensemble de textes réunis en vue d'une analyse, il apparaît que le chercheur ne travaille pas sur un corpus unique, mais sur une pluralité de corpus. Chacun d'eux remplit une fonction spécifique dans le processus de recherche, qu'il s'agisse de la collecte, de la sélection, de l'analyse ou de l'interprétation des données. Cette multiplicité souligne la complexité du travail empirique en linguistique et l'importance d'une construction méthodique des données.

2.1.Le corpus existant (ou corpus latent) : l'ensemble des textes auxquels on peut avoir accès, dont on peut disposer. C'est généralement une masse « informe », non systématique, mal définie, aux contours incertains. Il est difficile d'en avoir une vue globale. Cet existant dépend de conditions étrangères à l'étude, qui ne sont pas toutes connues ni maîtrisées.

2.2.Le corpus de référence **:** il est composé à partir du corpus existant, en adéquation avec l'objectif de travail ; il est clairement défini et équilibré. C'est lui qui fournit l'univers le plus large dans lequel chaque élément trouve sa valeur. Il constitue l'univers et fixe le point de vue de l'étude. Il représente le fond sur lequel on veut profiler les textes étudiés. Autrement dit, il matérialise le contexte, actif et virtuel, et acquiert là son statut linguistique d'unité de rang supérieur –la linguistique ne s'arrêtant ni à la phrase, ni même au texte (Rastier 1998, §III.2).

1.3.Le corpus d'étude : c'est l'ensemble des textes sur lesquels porte effectivement l'analyse, pour lesquels on attend des enseignements, des résultats. Le corpus d'étude n'est pas nécessairement une partie du corpus de référence, mais le corpus de référence doit pouvoir être considéré comme représentatif du corpus d'étude, pour l'aspect dont on veut rendre compte. Paradoxalement, le corpus d'étude peut être plus volumineux que le corpus de référence : ce qui est définitoire, ce n'est donc ni un rapport d'inclusion, ni un rapport de taille, mais la spécificité des rôles de chacun

***1.4.Le corpus distingué* :** c'est un groupe de textes du corpus d'étude que l'on veut caractériser dans leur cohésion d'ensemble, par rapport au reste du corpus d'étude. (un sous corpus)

Le contexte d'étude fournit un (*corpus de référence, corpus d'étude*). Le *fonds documentaire* désignerait alors l'ensemble des textes à disposition (de préférence à *corpus existant*) ; quant au *corpus distingué*, il correspond habituellement à ce que l'on appelle *sous-corpus*. Le choix ici de décliner quatre *corpus*, malgré leurs différences profondes, se justifie par l'intention de souligner les usages contrastés du mot *corpus*.

Dans le champ de la linguistique, les corpus occupent une place centrale en tant qu'outils empiriques permettant l'observation systématique des faits de langue. Un corpus se définit comme un ensemble structuré de données linguistiques authentiques, recueillies selon des critères précis en fonction des objectifs de recherche. Il constitue ainsi une base essentielle pour une approche inductive du langage, fondée sur l'analyse de données réelles plutôt que sur des intuitions ou des exemples construits.

Le rôle principal des corpus est de rendre possible une description fine et objective des usages linguistiques. En effet, ils permettent d'accéder à la langue telle qu'elle est effectivement produite dans des contextes variés (oral, écrit, numérique, institutionnel, etc.), favorisant ainsi une meilleure compréhension des régularités, des variations et des évolutions linguistiques. Cette dimension est particulièrement importante dans des domaines comme la sociolinguistique, la pragmatique ou l'analyse du discours, où l'attention est portée sur les usages situés et les pratiques langagières.

Par ailleurs, les corpus remplissent une fonction explicative en offrant aux chercheurs la possibilité de formuler et de tester des hypothèses. Grâce à des méthodes d'analyse quantitatives et qualitatives, ils permettent de mettre en évidence des tendances, des fréquences, des cooccurrences ou encore des structures récurrentes. Dans cette perspective, les corpus ne se limitent pas à une simple collection de données, mais deviennent de véritables instruments de validation scientifique, contribuant à renforcer la rigueur méthodologique des recherches en linguistique.

En outre, l'exploitation des corpus s'inscrit dans le développement des approches computationnelles et de la linguistique de corpus, qui mobilisent des outils informatiques pour traiter de grandes quantités de données. Ces avancées ont considérablement élargi les possibilités

d'analyse, notamment en permettant d'explorer des phénomènes linguistiques à grande échelle et de croiser différents niveaux d'analyse (lexical, syntaxique, discursif).

Enfin, les corpus jouent également un rôle didactique et appliqué. Ils sont utilisés dans l'enseignement des langues pour illustrer des usages authentiques, concevoir des activités pédagogiques ou encore élaborer des ressources lexicographiques et grammaticales. Ainsi, leur fonction dépasse le cadre strict de la recherche pour s'étendre à des domaines tels que la didactique des langues, la traduction ou encore le traitement automatique des langues.

CONCLUSION

Le corpus fournit à la fois des éléments à étudier, mais aussi l'environnement descriptif de ces éléments. Le corpus est un tout, un vaste ensemble, qui constitue à lui seul le cadre et le référentiel de l'analyse ; Il met en présence les éléments, et assure leur interrelation globale. Ces éléments prennent alors une valeur relative par rapport au corpus : affinités et associations, fréquence ou rareté, banalité ou spécificité, etc

En somme, les corpus constituent un outil incontournable pour l'observation, la description et l'explication des faits de langue, en offrant aux linguistes un accès direct aux données empiriques et en favorisant une approche scientifique fondée sur l'analyse des usages réels.

CHAPITRE II : CONSTITUTION ET EXPLOITATION D'UN CORPUS

- # CHAPITRE II : CONSTITUTION ET EXPLOITATION D'UN CORPUS

  Objectifs du chapitre

  Ce chapitre a pour finalité de familiariser les étudiants avec les procédures concrètes de constitution et de collecte de données linguistiques.

- Comprendre les critères fondamentaux de constitution d'un corpus (représentativité, homogénéité, pertinence).
- Apprendre à concevoir un corpus en fonction d'une problématique de recherche précise.
- Identifier les difficultés méthodologiques liées à la collecte des données linguistiques.
- Maîtriser les principaux outils de collecte des données (observation, entretien, questionnaire, analyse de contenu, expérimentation).
- Savoir recueillir et organiser des données orales et écrites accompagnées de métadonnées.
- Développer des compétences pratiques dans la transcription, le tri et la structuration des données.
- S'initier à une démarche rigoureuse de terrain en linguistique.

CHAPITRE II : CONSTITUTION ET EXPLOITATION D'UN CORPUS

COURS 06: CONSTITUER UN CORPUS : CRITERES DE RECUEIL.

INTRODUCTION

Le problème de la constitution d'un corpus et de recueil des données occupe une place centrale en domaine des sciences du langage en général et tout particulièrement dans le champ de l'analyse de discours depuis ses débuts, c'est-à-dire à la fin des années 1970. C'est à dire que la problématique de constitution, de collecte, de confection, de sélection – les termes ne manquent pas – des données d'analyse reste posée et qu'une question majeure, celle de voir quels sont (ou seraient) les critères fiables, précis, permettant à toute recherche d'asseoir les contours en fonction de son corpus.

Dans cette partie, on se focalise sur les critères susceptibles de marquer le recueil d'un corpus particulier.

- Critères de sélection (domaine, genre, époque, etc.)
- Outils, méthodes et étapes de collecte de données.
- Difficultés de constitutions de corpus.

Structurer, annoter et analyser un corpus (Etude de corpus authentiques)

1. CRITERES DE SELECTION DE CORPUS (PROPRIETES RECHERCHEES)

Choisir entre tel ou tel corpus dépend des objectifs visés :

- Etude de la grammaire
- Analyse terminologique
- Analyse conversationnelle, discursive.
- Etude lexicographique d'un dictionnaire, d'une variante du français
- Recherche de néologismes.
- Description des langues et des parlers.
- Etude des représentations des langues.
- Etude comparative des expressions de langues…etc

Le corpus nous le rappelons, est un ensemble de données à analyser en vue de répondre à une problématique. Le corpus ne se laisse pas uniquement définir formellement, comme un ensemble de textes. *Il vérifie* ***trois*** *types de conditions* ***: des conditions de signifiance, des conditions d'acceptabilité, et des conditions d'exploitabilité.***

• ***Conditions de signifiance*** : Un corpus est constitué en vue d'une étude déterminée (*pertinence*), portant sur un objet particulier, une réalité telle qu'elle est perçue sous un certain angle de vue (et non sur plusieurs thèmes ou facettes indépendants, simultanément) (*cohérence*).

• ***Conditions d'acceptabilité*** : Le corpus doit apporter une représentation fidèle (*représentativité*), sans être parasité par des contraintes externes (*régularité*). Il doit avoir une ampleur et un niveau de détail adaptés au degré de finesse et à la richesse attendue en résultat de l'analyse (*complétude*).

• ***Conditions d'exploitabilité*** : Les textes qui forment le corpus doivent être commensurables ou semblables (*homogénéité*). Le corpus doit apporter suffisamment d'éléments pour pouvoir repérer des comportements ou phénomènes significatifs (**au sens statistique du terme) (*volume*).**

Chacune de ces conditions demande à être commentée, à partir des éclairages complémentaires, et assez remarquablement convergents, issus des différentes disciplines qui utilisent les corpus.

1.1.Pertinence

Le corpus prend sens par rapport à **un objectif d'analyse**. Cela n'est pas sans incidence sur la question de sa réutilisabilité : à quelles conditions ce qui a été rassemblé pour servir un

objectif peut être recyclé pour en servir un autre ? Une partie de la réponse se trouve dans l'explicitation des choix et conditions de recueil du corpus. D'autre part, ce n'est pas nécessairement le corpus tel quel qui est repris : le corpus original (existant) sert de source pour construire un autre corpus (de référence), dans le respect du nouveau contexte d'analyse.

Règle de pertinence : Les documents retenus doivent être adéquats comme ***source*** *d'information pour correspondre à l'objectif qui suscite l'analyse*. (Bardin 1977, §III.I.1, p. 128)

[...] ce qui permet [par exemple] de réunir une cinquantaine de réponses individuelles en corpus collectif, c'est un ensemble de caractères communs aux testés : leur appartenance à la même communauté linguistique, à la même classe d'âge ; c'est aussi le même niveau culturel, la même« situation de testés ». (Greimas 1966, §VI.3, pp. 93-94)

1.2. Représentativité

Pour définir un échantillon représentatif il faut s'assurer que l'extrait présente la même configuration des observables. La réalité à décrire présente un certain équilibre, une certaine composition, que le corpus doit d'efforcer de refléter.

1 .3.Règle de représentativité : L'échantillonnage est dit rigoureux si l'échantillon est une partie **représentative** de l'univers de départ (ou corpus existant). Dans ce cas les résultats obtenus sur échantillon seront généralisables à tout l'ensemble.

Pour **échantillonner** il faut pouvoir repérer la distribution des caractères des éléments de l'échantillon**.** *Un univers hétérogène demande un échantillon plus important qu'un univers homogène*. [...] Comme pour un sondage, l'échantillonnage peut se faire au hasard, ou par quotas (les fréquences des caractéristiques de la population étant connues, on les reprend dans des populations réduites pour l'échantillon).(Bardin 1977, §III.I.1, p. 127)

Pour la linguistique, ce qui autorise des études sur des corpus toujours limités, c'est la nature redondante de la langue et la clôture des unités textuelles.

Le corpus n'est [...] jamais que partiel, et ce serait renoncer à la description que de chercher à assimiler, sans plus, l'idée de sa représentativité à celle de la totalité de la manifestation. Ce qui permet de soutenir que le corpus, tout en restant partiel, peut être représentatif, ce sont les traits fondamentaux du fonctionnement du discours retenus sous les noms de redondance et de clôture. Nous avons vu que toute manifestation est itérative, que le

discours tend très vite à se fermer sur lui-même : autrement dit, la manière d'être du discours porte en elle même les conditions de sa représentativité. (Greimas 1966, §IX.1.b, p. 143)

Quand l'étude vise à décrire la langue ou le fonctionnement des textes « en général », la condition de représentativité semble devoir se traduire par une recherche de diversité maximale. Autrement dit, dans l'idéal, tous les cas de figure existants doivent être présents dans le corpus. Deux tactiques sont observables : la course à la quantité d'une part (engranger le maximum de données, le poids total devant être garant de la richesse amassée), la construction raisonnée d'autre part (se donner une grille quadrillant la réalité, et s'en servir pour rassembler méthodiquement des textes correspondant à tous les aspects recensés). La première tactique, dont la devise est « more data is better data » (Péry-Woodley 1995, §2.3.1), est manifestement grossière, mais souvent elle est justifiée (en partie) par les difficultés profondes auxquelles se heurte de plein fouet la seconde tactique : quel modèle adopter pour organiser la sélection des textes, qui ne porte pas sa part d'a priori réducteurs ? Plus gravement, la problématique elle-même apparaît utopique irréaliste : il n'y a pas de langue générale, ou standard, ou moyenne ; et les textes sont tous pris dans des pratiques qui les contextualisent.

La recherche de corpus équilibrés semble bien constituer une impasse : la notion d'équilibre s'apparente à celle de « langue générale », et elle paraît tout aussi insaisissable. Elle suppose également une recherche irréaliste d'exhaustivité : le corpus équilibré est sans doute celui qui a « de tout un peu », mais encore faudrait-il savoir ce qu'est « tout », c'est-à-dire quelles sont les classes à représenter, –ce qui nécessite un modèle complet de la variation –, et avoir accès à des textes les représentant. (Péry-Woodley 1995, §2.3.2, p. 218)

Admettre la relativité et la part de choix qu'il y a dans la constitution de tout corpus, c'est également reconnaître le caractère décisif de l'établissement du corpus. En particulier, bien souvent le corpus (ou une de ses parties) est utilisé comme référentiel (puisqu'il est représentatif de la réalité à décrire) et il conditionne tous les résultats de l'analyse.

1.4.Régularité et exhaustivité

La régularité correspond au fait que l'on explicite des principes pour définir le corpus, sans se permettre d'exceptions qui introduiraient des écarts locaux (manques, excès, éléments étrangers).

L'exhaustivité du corpus est « à concevoir comme l'adéquation du modèle à construire à la totalité de ses éléments implicitement contenus dans le corpus. (Greimas 1966, §IX.1.b, p. 143)

Règle ou exigence de l'exhaustivité : une fois défini le champ du corpus (entretiens d'une enquête, réponses à un questionnaire, éditoriaux d'un quotidien de Paris entre telle et telle date, émissions de télévision concernant tel sujet, etc.), il faut prendre en compte tous les éléments de celui-ci. Autrement dit, il n'y a pas lieu de laisser un élément pour une raison quelconque (difficulté d'accès, impression de non-intérêt) non justifiable sur le plan de la rigueur.

Par exemple, on réunit un matériel d'analyse des publicités pour automobiles parues dans la presse pendant une année. Toute annonce publicitaire répondant à ces critères doit être recensée. (Bardin 1977, §III.I.1, p. 127)

L'exigence d'exhaustivité concerne les ensembles des individus et des variables représentant un inventaire complet d'un domaine réel .

En règle d'exhaustivité, le corpus doit avoir aussi un niveau de détail adapté aux besoins de l'analyse : les adaptations nécessaires peuvent être soit de l'enrichir et de l'affiner, soit d'ajuster, par réduction, le niveau de discrétisation de la réalité à représenter réalisée à partir des données. L'exhaustivité des données peut,conformément au principe d'équivalence distributionnelle, être assurée par une partition [...], ou [par le] choix d'un échantillon fini (éventuellement stratifié [...]) sur un espace potentiel continu (Benzécri & al. 1973, § Indice systématique)

1.5.Homogénéité

Sachant l'objectif de l'analyse, et les dimensions de variation que l'on veut étudier, le corpus doit être aussi homogène que possible pour ses autres caractéristiques.

Exigence d'homogénéité : toutes les données recensées [...] sont des quantités de même nature. (Benzécri & al. 1973b, §A.2.1.3, p. 21)

Pour définir objectivement des données étudiées, on vise **l'homogénéité des variables** : ce qui permet l'adoption d'une unité de mesure unique.

En règle ou exigence d'homogénéité, les documents retenus doivent être homogènes, c'est-à-dire obéir à des critères *de choix précis et ne pas présenter trop de singularité en dehors de ces critères de choix.*Par exemple, des entretiens d'enquête, effectués sur un thème donné, doivent : être tous concernés par ce thème, avoir été obtenus par des techniques

identiques, être le fait d'individus comparables. Cette règle est surtout utilisée lorsqu'on désire obtenir des résultats globaux ou comparer les résultats individuels entre eux. (Bardin 1977, §III.I.1, p. 128)

Lorsque nous utilisons le terme corpus, nous sous-entendons 'corpus de documents homogènes', à savoir un ensemble de documents qui ne soit pas hétéroclite. Il ne s'agit pas de considérer n'importe quel ensemble de documents sans aucun rapport les uns avec les autres. Par exemple, un ensemble de publications mondiales sur l'intelligence artificielle constituent pour nous, des corpus homogènes. Les traitements que nous exposerons par la suite sont envisagés sur de tels corpus. (Chartron 1988, §II.1, p. 16)

Le choix d'un corpus présuppose que ce corpus constitue bien un objet d'étude ; c'est-à-dire, que l'analyste le perçoive comme une entité ou un objet dans l'univers référentiel qui l'intéresse. En définitive, même si ce n'est que de manière implicite, l'analyste fait des hypothèses sur les conditions d'existence de cet objet, sur ses lois de production, sur les paramètres qui le font reconnaître dans cet univers référentiel. (Reinert 1990, §1.2, p. 27)

1.6.Volume

Certains corpus de taille importante, ou ce qu'on appelle métadonnées, ou méga-corpus nécessitent de faire appel à un traitement statistique, à l'aide de logiciels particuliers. (Iramuteq, alceste…etc)

Les procédés d'analyse visent à saisir et décrire des régularités qui structurent le corpus. Une certaine redondance est nécessaire pour que puissent émerger et être repérés des aspects caractéristiques et informatifs.

Le logiciel ALCESTE est un outil d'aide à l'interprétation d'un corpus textuel : entretiens, réponses à une question ouverte, corpus littéraires, corpus numérique, en fait tout document textuel pourvu qu'il présente une certaine homogénéité et un volume important. [...]

Il y a toutefois deux conditions pour obtenir un résultat signifiant : la première est que le corpus présente une certaine cohérence thématique [condition d'homogénéité]. C'est le cas en général des réponses à une question ouverte, de textes littéraires, de recueils d'articles sur un sujet, etc...

La seconde est que le document soit **suffisamment volumineux** pour que l'élément statistique entre en ligne de compte. C'est du reste l'intérêt d'ALCESTE de donner très rapidement une vision globale sur une documentation volumineuse qui serait autrement très longue à dépouiller. (Reinert, Piat 1995, cahier 1, §0, p.3)

La condition de volume est importante pour des analyses statistiques, pour que celles-ci puissent être considérées significatives. En revanche, présenter la recherche de volume essentiellement comme un moyen d'obtenir une bonne représentativité 'générale' est **déplacé** : le volume et la représentativité sont des caractéristiques à part entière, complémentaires.

Dans le cas d'une exploitation manuelle, c'est-à-dire sans l'outil informatique, on s'inquiétera à l'inverse de la maniabilité du corpus.

CONCLUSION

En définitive, la constitution d'un corpus apparaît comme une étape déterminante et structurante de toute recherche en linguistique. Loin d'être une simple opération technique de collecte, elle engage des choix méthodologiques fondamentaux liés aux objectifs de l'analyse, aux critères de sélection et aux conditions de validité scientifique. Entre exigences de pertinence, de représentativité, d'homogénéité et d'exploitabilité, le chercheur doit constamment trouver un équilibre entre rigueur théorique et contraintes empiriques.

Par ailleurs, les difficultés inhérentes à la recherche d'un corpus « idéal » – notamment en matière de représentativité ou d'exhaustivité – rappellent que tout corpus est une construction située, partielle et orientée. C'est précisément cette dimension construite qui exige une explicitation rigoureuse des choix opérés. Ainsi, loin d'être un simple support, le corpus constitue un véritable objet scientifique qui conditionne en grande partie les résultats et les interprétations de l'analyse.

COURS07 : DIFFICULTES DE CONSTITUTION DE CORPUS.

INTRODUCTION

La constitution d'un corpus en linguistique soulève un ensemble de difficultés méthodologiques, dont la plus déterminante concerne la notion de représentativité. En effet, un corpus est considéré comme « fiable » ou « pertinent » dans la mesure où il est capable de refléter, à une échelle réduite, les caractéristiques essentielles d'un ensemble discursif plus large. Autrement dit, il doit contenir un échantillon suffisamment équilibré et significatif pour permettre de généraliser les résultats de l'analyse à d'autres productions du même type. Cette exigence implique un travail rigoureux de sélection et d'échantillonnage, qui ne va pas sans poser de nombreux défis, notamment lorsqu'il s'agit de rendre compte de la diversité des usages linguistiques.

1. DIFFICULTES DE CONSTITUTIONS DES CORPUS.

Il semblerait bien que la difficulté d'assembler ces données d'analyse réside dans la notion de représentativité. En effet, un corpus « fiable », « sûr » est un corpus représentatif ; autrement dit qu'il suffit à lui seul pour rendre compte d'un ensemble de critères et de traits pouvant être transposés sur tous les autres discours (du même type).

Cela revient à dire donc que cette notion de représentativité assure le lien qui ramène le corpus d'étude « réduit » au corpus général de référence, sans oublier la question de l'authenticité des données sélectionnées pour l'analyse. Le corpus sélectionné doit être susceptible d'apporter un élément nouveau et bouleverser les résultats, à moins de restreindre dès le départ, les visées et objectifs de l'analyse.

La représentativité constitue ainsi le principe qui assure le passage du corpus d'étude – nécessairement limité – à un corpus de référence plus global. Toutefois, cette opération reste délicate, car elle suppose de déterminer quels critères doivent être retenus pour que l'échantillon reflète fidèlement la réalité étudiée. À cette difficulté s'ajoute la question de l'authenticité des données : les textes sélectionnés doivent être attestés, situés dans des contextes réels d'usage, et non fabriqués artificiellement. Par ailleurs, un corpus pertinent doit être en mesure d'apporter des éléments nouveaux à l'analyse, voire de remettre en question certaines hypothèses initiales, à moins que le chercheur ne choisisse volontairement de restreindre ses objectifs dès le départ.

Le choix d'un corpus dépend donc étroitement des objectifs de recherche poursuivis. Ces objectifs peuvent être multiples : il peut s'agir, par exemple, d'une étude grammaticale visant à décrire des structures syntaxiques, d'une analyse terminologique centrée sur un domaine spécialisé, ou encore d'une analyse conversationnelle ou discursive portant sur les interactions langagières. Les corpus peuvent également servir à des travaux lexicographiques (élaboration de dictionnaires ou étude de variantes linguistiques), à l'identification de néologismes, à la description des langues et des parlers, à l'étude des représentations sociales des langues ou encore à des comparaisons interlinguistiques. Ainsi, la diversité des finalités implique des choix méthodologiques différenciés dans la constitution des corpus.

Au-delà de ces objectifs, un corpus ne peut être réduit à une simple collection de textes. Il doit répondre à un ensemble de conditions qui garantissent sa validité scientifique. On distingue généralement trois grandes catégories de critères : les conditions de signifiance, les conditions d'acceptabilité et les conditions d'exploitabilité.

Les conditions de signifiance renvoient à la pertinence et à la cohérence du corpus. Celui-ci doit être constitué en fonction d'un objectif précis et porter sur un objet d'étude clairement défini. Il ne s'agit pas d'accumuler des données hétérogènes, mais de sélectionner des textes qui éclairent une problématique spécifique. La pertinence implique que les documents retenus soient adéquats par rapport à la question de recherche, tandis que la cohérence suppose une certaine unité dans le point de vue adopté. Comme le souligne Bardin, les données doivent correspondre directement à l'objectif de l'analyse, ce qui implique une sélection raisonnée. De même, les travaux de Algirdas Julien Greimas montrent que la constitution d'un corpus repose souvent sur l'identification de traits communs entre les données (appartenance sociale, âge, contexte, etc.), garantissant ainsi une base d'analyse homogène.

Les conditions d'acceptabilité concernent principalement la représentativité, la régularité et la complétude du corpus. Un corpus acceptable doit offrir une image fidèle de la réalité étudiée, sans être biaisé par des contraintes externes. La représentativité repose sur un échantillonnage rigoureux : l'échantillon doit reproduire la distribution des caractéristiques observées dans l'ensemble de référence. Comme dans les enquêtes statistiques, cet échantillonnage peut être aléatoire ou construit par quotas. Toutefois, en linguistique, la question est complexifiée par la nature même du langage, qui est à la fois varié et contextuel.

La complétude, quant à elle, implique que le corpus soit suffisamment riche et détaillé pour permettre une analyse fine, tandis que la régularité suppose une certaine stabilité dans les conditions de production des données.

Les conditions d'exploitabilité renvoient à la possibilité d'analyser efficacement le corpus. Cela suppose que les données soient comparables entre elles (homogénéité) et suffisamment nombreuses pour permettre l'identification de phénomènes significatifs (volume). Un corpus trop restreint ou trop hétérogène risque de limiter la portée des résultats. À l'inverse, un corpus bien construit offre un terrain propice à des analyses à la fois qualitatives et quantitatives.

Enfin, la question de la représentativité en linguistique doit être nuancée. Comme le souligne Algirdas Julien Greimas, un corpus n'est jamais exhaustif, mais il peut néanmoins être représentatif grâce à deux huuոlų fondamentales du langage : la redondance et la clôture des discours. Les phénomènes linguistiques tendent à se répéter, ce qui permet d'observer des régularités même à partir d'un corpus limité. Dès lors, la représentativité ne signifie pas exhaustivité, mais capacité à rendre compte des mécanismes essentiels du fonctionnement discursif.

Dans les études visant une description générale de la langue, la représentativité tend à être recherchée à travers la diversité maximale des données. Deux approches peuvent alors être envisagées : soit accumuler un grand volume de données (approche quantitative), soit construire un corpus de manière raisonnée en s'appuyant sur une grille de sélection (approche qualitative). Toutefois, chacune de ces stratégies présente des limites, notamment en raison du caractère fondamentalement hétérogène et contextualisé des pratiques langagières.

CONCLUSION

Il apparaît ainsi que la constitution d'un corpus relève d'un compromis entre ambition scientifique, contraintes méthodologiques et faisabilité empirique.

COURS 08: PRINCIPAUX OUTILS DE COLLECTE DE DONNEES : L'OBSERVATION

Le corpus est généralement l'apanage d'une linguistique descriptive, qui l'observe pour reconstituer à posteriori des régularités.

INTRODUCTION

Dans le cadre des recherches en linguistique, et plus particulièrement dans les approches empiriques, la collecte des données constitue une étape essentielle qui conditionne la qualité et la validité des analyses. Le choix des outils de recueil ne se fait pas de manière arbitraire : il dépend étroitement de la problématique de recherche, des hypothèses formulées ainsi que de la nature des phénomènes à observer. Autrement dit, toute démarche scientifique suppose une adéquation entre les objectifs poursuivis et les méthodes mobilisées pour accéder aux données.

Parmi les différents dispositifs méthodologiques disponibles, l'observation occupe une place privilégiée. Elle permet d'accéder directement aux pratiques langagières dans leur contexte naturel, en offrant au chercheur la possibilité de saisir les usages tels qu'ils se manifestent réellement. À ce titre, elle constitue un outil fondamental pour l'étude des interactions, des comportements discursifs et des situations de communication.

Dans cette perspective, il convient de présenter les principales techniques de collecte de données, en mettant l'accent sur l'observation en situation, ses modalités de mise en œuvre et les questions méthodologiques qu'elle soulève.

1. L'OBSERVATION OU (L'observation en situation)

1.1.DEFINITION

L'observation se définit comme étant le lieu ou le contexte institutionnel dans lequel se confronte le modèle d'analyse à des données observables. Au cours de ce processus, de nombreuses informations (réponses à des questionnaires, données statistiques,

observations, documents, etc.) sont rassemblées afin d'être exploitées dans une étape ultérieure.

La mise en œuvre de cette étape d'observation nous conduit à répondre à trois questions principales :

-Observer quoi ? : la réponse consiste à définir des données pertinentes Il faut d'abord commencer par déterminer un objet d'étude, puis après quelques séquences d'observations globales, arriver à identifier la nature du phénomène observé et les interrogations qu'on se pose. Il s'agit d'accéder à toutes les données pertinentes, c'est-à-dire, celles qui contribuent efficacement à la vérification des hypothèses.

1.2.Caractéristiques de l'observation

Observer qui ?A ce stade il est conseillé de circonscrire de manière opérationnelle le contexte dans lequel l'étude se déroulera et les conditions de sa réalisation : le délai, le mode de sélection adaptée. Selon le cas, il s'agira d'étudier soit la totalité de la communauté considérée, soit seulement un échantillon représentatif ou caractéristique de cette communauté.

-Comment observer ? il s'agit de choisir les instruments d'observation et de collecte de données. L'observation est considérée comme un moyen d'accès au terrain, à l'information et se distingue aussi d'autres formes de recueil d'informations.

Pour Grawitz, « *l'observation sur le terrain pose un grand nombre de problèmes en fonction de l'objectif que l'on vise et de la situation devant laquelle on se trouve. La taille du groupement, la complexité des interactions, la précision des objectifs, impliquent le choix de techniques adaptées* » (1993 :53).

Cette technique s'appuie sur l'aspect visuel et consiste, contrairement aux autres techniques, à recueillir les données qualitatives et quantitatives d'une manière directe sans recourir aux témoignages ou aux documents. Elle étudie les attitudes et les comportements des individus ou des groupes, en prenant en considération l'aspect idéologique, culturel et même social.

Le chercheur canalise son observation en fonction de ses objectifs et de ses hypothèses du travail.

L'objectif principal de cette technique est d'étudier tout ce qui relève du non verbal ainsi que des comportements définis, des profils de comportements et des modes de vie tels

qu'ils se sont produit. Pour ce faire, le chercheur doit être muni d'une grille d'observation établie en fonction des comportements à observer.

2.3. Types d'observation

Il existe plusieurs types d'observations dont les plus usuelles **:**

- **L'observation participante :** elle consiste à faire partie de la vie collective de la communauté étudiée afin de pouvoir établir des constats. Elle peut se réaliser aussi en interagissant avec les personnes observées pour poser des questions. Le chercheur fait appel à cette technique quand il s'agit d'obtenir des informations détaillées sur une culture particulière, un groupe de personne ou sur le comportement des individus.
- **L'observation non participante :** elle consiste à observer de l'extérieur la communauté visée, en demeurant en retrait de la vie des sujets. Ce type d'observation est utilisé dans le cas où le chercheur tend à préserver la situation observée dans son état naturel, original, pour augmenter la fiabilité des résultats. Cette méthode permet d'avoir une idée générale, un aperçu sur une situation non étudiée jusque là et de définir le sujet de recherche pour déterminer la technique adéquate d'analyse.

Quand utiliser l'observation en situation ?

- Quand il s'agit d'expliquer le fonctionnement d'un phénomène.
- Quand il s'agit d'étudier un phénomène dans son cadre naturel.
- Quand il s'agit d'un sujet peu connu.
- Quand il s'agit d'avoir accès à la réalité (avec exactitude) sur une situation.

Comment préparer une observation : pour mener à bien son observation le chercheur soit suivre un protocole qui consiste à :

- *Définir le sujet à observer* : il faut déterminer la cible ainsi que son contexte.
- *Choisir le type d'observation* : parmi les quatre types proposés il faut choisir le plus fiable pour recueillir des données pertinentes.

- *Préparer le protocole de l'observation* : voir le terrain de l'observation, la durée, les personnes concernées et les outils facilitant la tâche.
- *Repérer le lieu d'observation* : visiter le lieu d'observation avant le jour J pour étudier en toute discrétion l'emplacement de l'enquêteur, la taille du lieu, etc.
- *Réaliser un schéma d'observation* : si le chercheur opte pour une observation systématique, ce schéma est obligatoire. Dans les autres cas il est conseillé de le préparer pour structurer l'observation.
- *Préparer le matériel* : smartphone, schéma d'observation, carnet, le chercheur doit préparer son matériel pour réussir son observation.

2.Comment réaliser une observation ?

Pour une observation participante, l'enquêteur se présente rapidement et parle brièvement de son travail pour ne pas influencer le comportement des sujets observés.

-Se faire le plus discret possible pour observer la réalité telle qu'elle est et ne pas inciter le sujet à changer ses comportements.

-Adopter une démarche neutre pour effectuer une observation fiable des faits, des gestes, et des individus observés.

- Utiliser (dans la mesure du possible) le smartphone : pour garantir une totale discrétion, l'observateur utilise son smartphone pour prendre ses notes sans dévoiler son attitude d'observateur ou susciter des interrogations sur son attitude.

-Procéder à l'analyse : l'observateur relit ses notes après son observation, détermine les éléments notables, résume les idées principales pour être en mesure de rédiger sa conclusion.

3. Avantages et limites de l'observation

L'observation constitue un outil méthodologique particulièrement intéressant en linguistique, notamment en raison de la richesse et de la profondeur des informations qu'elle permet de recueillir. En donnant accès à des données produites dans des situations réelles, elle favorise une relative spontanéité du matériau linguistique, ce qui renforce la validité des analyses. Par ailleurs, l'observation offre au chercheur la possibilité d'intervenir de manière ponctuelle, en posant des questions ou en sollicitant des précisions afin d'obtenir des informations complémentaires. Elle se révèle également utile dans les phases exploratoires de la recherche, notamment pour définir un sujet d'étude ou poser les bases d'une

problématique. Enfin, elle permet d'appréhender globalement une situation ou un comportement, en offrant une vision d'ensemble des phénomènes observés.

Cependant, malgré ses nombreux atouts, l'observation présente aussi certaines limites qu'il convient de prendre en compte. L'une des principales difficultés réside dans le risque de subjectivité de l'observateur, dont les perceptions et les interprétations peuvent influencer l'analyse. Ce biais peut conduire à des lectures erronées ou insuffisamment fondées scientifiquement des comportements observés. En outre, la présence du chercheur peut modifier le comportement des individus observés, qui tendent parfois à adopter des attitudes différentes de leurs pratiques habituelles, ce qui affecte l'authenticité des données recueillies. Enfin, l'observation ne garantit pas l'exhaustivité des informations : certains éléments peuvent échapper à l'attention du chercheur, entraînant ainsi des lacunes dans l'analyse.

Avantages	**Limites**
- richesse, profondeur de l'information recueillie, - spontanéité relative du matériau, - possibilité d'apporter un questionnement pour délivrer des informations supplémentaires, - utilité pour définir un sujet de recherche et constituer la base d'une recherche, - possibilité d'avoir une vue générale de la situation ou du comportement à étudier.	- Risque De subjectivité de l'observateur. - Risque d'interprétations erronées et non scientifiques des résultats et des comportements observés. - Risque de détour du sujet étudié : il adopte un comportement différent de son comportement habituel. - Risque de passer à côté de certaines informations.

4.Le recueil des données existantes :

Les données secondaires et données documentaires : C'est rassembler des documents en vue de les étudier tels qu'ils sont ou pour servir à l'étude d'un objet intéressant. Elles permettent de mettre en lumière des phénomènes que l'enquêteur ne parvient pas à recueillir à leur propos, d'une manière directe où dans le cas ou les données recueillies sont insuffisantes. Par exemple étudier : les phénomènes linguistiques via les réseaux sociaux, macrosociaux, démographiques, les idéologies et le développement historique des phénomènes sociaux. Il faut souligner que ce modèle de recueil des données s'avère très utile dans la mesure où il permet un gain du temps, d'argent et, en conséquence, dévoile l'utilité des martiaux documentaires dans le domaine de la recherche où les outils technologique règnent.

Au final, il faut souligner que rien n'empêche un chercheur de rassembler des données en utilisant plus d'un instrument de collecte. Au contraire une telle pratique lui permet de mieux valider ses données. En effet, si deux sources d'informations révèlent les mêmes faits, le chercheur sera convaincu de la validité des informations ainsi recueillies.

CONCLUSION

En somme, l'observation s'impose comme un outil fondamental de collecte des données en linguistique, en ce qu'elle permet d'accéder aux pratiques langagières dans leur contexte réel d'usage. Elle offre au chercheur une compréhension fine et contextualisée des phénomènes étudiés, tout en constituant une base solide pour l'élaboration et la validation des hypothèses. Toutefois, son efficacité dépend étroitement de la rigueur méthodologique mise en œuvre et de la capacité à limiter les biais liés à la subjectivité ou aux conditions d'observation. Ainsi, bien maîtrisée, l'observation demeure un levier essentiel pour produire des analyses pertinentes et scientifiquement fondées.

COURS 09 : OUTILS ET TECHNIQUES DE COLLECTE DES DONNEES : LE QUESTIONNAIRE.

(L'ENQUETE PAR QUESTIONNAIRE)

Introduction

Le questionnaire est un outil indispensable dans les travaux de recherches en linguistique centrés sur l'urbanité langagière, les représentations, les perceptions d'une situation linguistique ainsi que les pratiques linguistiques. Ainsi, la collecte d'informations via un questionnaire permet de saisir des renseignements utiles sur l'origine, le sexe, l'âge, les pratiques langagières, dégager les représentations que les habitants d'une ville se font de leur milieu urbain par exemple.

Contrairement à l'entretien et à l'observation qui sont des outils d'enquête individuels ou collectifs, **le questionnaire est uniquement collectif. (Outil d'enquête quantitatif)**

2. Le questionnaire : éléments de définition

C'est un outil méthodologique qui comporte un ensemble de questions structuré qui peut être conçu sous format : papier et électronique au près d'une population cible (groupe) et qui vise à comprendre et expliquer des faits linguistiques : « ***Le questionnaire est un outil qui permet de prélever des réponses de manière systématique*** **» (Anger, 1996 :16).**

C'est la technique la plus répondue après ***l'interview*** et qui se différencie du *sondage d'opinion* par son objectif visant la vérification d'hypothèses.

Les données obtenues peuvent figurer sous forme de graphiques quand elles sont quantitativement importantes (200 questionnaires)...

> « *Le sociolinguiste élabore le questionnaire dans le but de confronter avec les données empiriques , la pertinence des questions qu'il se propose d'élucider et de confirmer la validité des hypothèses postulées dans la phase préliminaire de sa recherche »*

Selon Combessie, le questionnaire est un instrument d'enquête permettant d'élargir le champ d'investigation et de tester la validité des hypothèses : il a pour fonction de « vérifier statistiquement jusqu'à quel point sont généralisables les informations et hypothèses préalablement constituées » .

Dans une perspective sociologique, le questionnaire est défini comme un outil central de construction du terrain. Il s'agit d'un instrument permettant de recueillir des données sur des pratiques, des représentations ou des comportements, en vue de construire scientifiquement un objet

d'étude .

Par ailleurs, certains auteurs insistent sur son caractère standardisé. Le questionnaire est ainsi considéré comme *« une batterie de questions standardisées » permettant de coder et de classer les informations recueillies de manière systématique, afin de constituer une base empirique exploitable scientifiquement* .

Dans une définition plus générale, le questionnaire peut être envisagé comme un instrument de recherche composé d'une série de questions ordonnées visant à recueillir des informations auprès d'un ensemble d'individus dans le cadre d'une enquête ou d'une étude statistique .

Enfin, dans les travaux de méthodologie en sciences sociales, il est également défini comme une technique de collecte de données quantifiables reposant sur un ensemble structuré de questions posées de manière identique à un grand nombre de répondants, afin de permettre des comparaisons et des analyses statistiques

2.Types de questionnaires

Au niveau de la forme et du contenu , on distingue trois types :

- **Questionnaire ouvert** : donne aux répondants la liberté de fournir des réponses à longueur variable (sur le plan de forme et de contenu). Celui-ci s'apparente à l'entretien directif qui vise à obtenir des opinions.
- **Questionnaire fermé** : ou question à choix multiple dont la formulation contient une liste préétablie de réponses possibles. Les questions proposées contraignent l'informateur à des choix de réponses limitées (vise à obtrenir des données factuelles/ sur des faits)
- **Questionnaire structuré** :qui combine les deux types précédents. Il vise à recueillir des données factuelles (concernent les faits, autrement dit des éléments objectifs, observables et faciles à déterminer :âge, sexe, profession) et des données sur des opinions.

Nous retenons d'autres types essentiels de questions qu'on peut introduire dans un questionnaire :

1.Les questions ouvertes d'opinions –qu'on appelle questions psychologiques- qui impliquent des réponses libres.

2.Les questions fermées pour les données factuelles : cocher la bonne réponse. Elles sont :

-La question avec échelle de valeur ou d'évaluation : on demande de répondre par un score (pour mesurer la préférences, ou le degré de satisfaction). Ou

-A échelle d'appréciation : un avis sur un sujet.

-La question dichotomique : **à choix binaire**, permet le choix entre deux réponses (vrai ou faux ; oui ou non)

3.Critères d'élaboration d'un questionnaire

Tout questionnaire est construit et conçu en fonction, d'une part, *des objectifs de la recherche* Et d'autre part, *des éléments contenus dans les hypothèses retenues* que le questionnaire doit interroger à travers ses questionnements.

Pour mener à bien l'élaboration d'un questionnaire et s'assurer de sa cohérence, de sa clarté, il faut s'assurer des éléments formels suivants :

- **La clarté des questions** : la question posée doit avoir un seul sens et une seule interprétation (emploi d'une terminologie simple, précise et univoque).
- **la concision du vocabulaire de la formulation** : élaborer les questions qui relèvent du langage courant et adapté de l'enquêté. Il ne faut pas utiliser les termes techniques ou le jargon de spécialité quel que soit le degré d'instruction de l'enquêté.
- **La neutralité dans la construction** : il ne faut pas intégrer dans les questions subjectives ou des éléments directifs sur la réponse.
- **La conformité de la question au contexte sociolinguistique** : il ne faut pas poser des questions qui ne sont pas conformes aux contextes culturels, idéologiques, et qui induisent au conflit.
- **Le nombre réduit des questions** : le questionnaire ne doit pas contenir beaucoup de questions (entraîne de la lassitude) ni trop réduit (accusé de légèreté).
- **L'organisation progressive (regrouper les questions par thème)** : l'élaboration du questionnaire doit suivre un certain enchaînement qui part du général au particulier, du plus simple ou plus complexe. Chaque partie doit correspondre à un **thème précis** et avoir un le même nombre proportion de questions posées (fermées et ouvertes).

Sur le plan du contenu :

-S'assurer de la pertinence et de l'utilité des questions : quelle information dois-je obtenir ? -De combien de question ai-je besoin pour obtenir cette information ?

L'enquêté est-il en mesure de répondre et de fournir réellement cette information pertinente ?

4.La structure d'un questionnaire

La présentation d'un questionnaire est un aspect fondamental, de ce fait il est recommandé de :

- **Mise en forme soignée du questionnaire** : respecter les espacements, les alignements et la clarté dans la mise en page. - Ne pas conserver la même forme de question pour ne pas entrainer de l'ennui, de la monotonie et surtout de l'automatisme.- Respecter les mêmes formats, dispositions et alignements des cases de réponses, pour faciliter la tâche aux répondants.
- **Mentionner toutes les informations** : permettant d'identifier et localiser les répondants.
- **Systématicité du questionnaire** : le même questionnaire doit être administré à tous les enquêtés.
- **Durée ou longueur de questionnaire** : le questionnaire doit contenir suffisamment de questions mais pas trop pour que les informateurs répondent vite.
- **Formuler des questions qui font sens / en relation avec le thème de recherche** : ce sont les questions relatives au sujet de la recherche.
- **Organisation et cohérence** : Commencer par les questions de fait, ensuite de fond (les plus brèves et les plus fermés placées en premier).
- **La neutralité des questions** : éviter les termes véhiculant les préjugés, la subjectivité.
- **L'exhaustivité des réponses suggérées** : Ajouter rubrique- autre- dans les questions à choix multiples.
- **Validation du questionnaire par un prétest** : pour s'assurer que les questions et la thématique soulevées sont cernées par les enquétés.

4.1.modalites d'élaboration de questionnaire

En effet, établir un questionnaire fiable implique la prise en considération de plusieurs éléments portant sur l'élaboration des questions qui doivent obéir à des **règles rigoureuses** afin d'obtenir les données escomptées en termes **de fiabilité et d'exploitation**.

Il y a deux façons d'administrer un questionnaire :

- Le chercheur peut rassembler les enquêtés et leur administrer lui- même le questionnaire. Il pourrait leur expliquer la question si l'enquêté ne la comprenait pas.

Le cas des élèves d'une classe, les employés d'une entreprise, les enseignants d'une école. Il s'agit d'un ***questionnaire d'administration directe***. Ce genre de questionnaire présente de nombreuses caractéristiques :

- ✓ Temps de réponse court et (bien géré);
- ✓ Connaissance de la population cible(pratiques, comportements, attitudes) ;
- ✓ Quantification d'une panoplie de données et analyse par croisement, recoupement ou corrélation ;
- ✓ Clarté des questions, vu la disponibilité du chercheur
- ✓ Taux élevé des participants ce qui permet d'étudier un phénomène social à travers un cas de représentativité.

Comme toute technique d'investigation, ce genre de questionnaire représente certaines limites et inconvénients qui se traduisent en termes *de la fiabilité de la technique (réponses superficielles, descriptions).*

- ➢ Le second type de questionnaire est ***d'administration indirecte*** où l'enquêteur se chargera de remplir le document pour l'enquêté. Cet instrument permet de recueillir une plus grande quantité d'informations. Il se caractérise par :
  - ✓ la quantité d'informations recueillies : il peut être envoyé à un très grand nombre de sujets sur un vaste territoire,
  - ✓ l'anonymat préservé .
  - ✓ le recueil rapide de l'information.

En revanche, ce genre de questionnaire pourrait être confronté à un certain nombre d'entraves à savoir :

-*Temps de réponse long*

- *Faiblesse du taux de réponse : ce qui sèmera le doute sur la représentativité du groupe des répondants.*

- Le répondant n'est pas toujours la personne concernée par le questionnaire.

CONCLUSION

En guise de synthèse, il est indispensable de souligner que dans une enquête par questionnaire on doit :

-Identifier nos objectifs.

-Fixer ce qu'on veut réaliser, proposer des hypothèses par exemple.

- Choisir la population ou les personnes concernées.

- Il faut choisir le nombre, le type et la qualité des questions.

- On doit les classer en partant du général au particulier.

- Exploiter et analyser les résultats: les décrire les résultats, les expliquer, répondre à la question de départ.

COURS 10: OUTILS ET TECHNIQUES DE COLLECTE DES DONNEES :L'ENTRETIEN (OU L'ENTREVUE)

INTRODUCTION

Dans les recherches en linguistique et plus largement en sciences humaines et sociales, le choix des outils de collecte des données dépend étroitement des objectifs poursuivis et de la nature des phénomènes étudiés. Parmi ces outils, l'entretien (ou entrevue) occupe une place privilégiée, notamment dans les approches qualitatives. Il permet d'accéder à des données riches, nuancées et contextualisées, en donnant la parole aux acteurs eux-mêmes.

Contrairement aux méthodes standardisées comme le questionnaire, l'entretien repose sur une interaction directe entre le chercheur et l'enquêté, favorisant l'expression des opinions, des expériences et des représentations. Il s'agit ainsi d'une situation de communication orientée vers un objectif de recherche, où l'enquêteur cherche à comprendre le sens que les individus attribuent à leurs pratiques et à leurs vécus . Dans cette perspective, l'entretien apparaît comme un outil essentiel pour explorer la dimension subjective et interprétative des faits linguistiques et sociaux.

1. Définition

Cette technique consiste à interroger d'une manière directe (contrairement au questionnaire) un groupe ou un individu afin de récolter des informations très riches et nuancées sur les attitudes, les comportements et les représentations de certains faits pour la population interrogée. Selon Quivy et L. Van Campenhoud, cette technique permet *« l'analyse du sens que les acteurs donnent à leurs pratiques et aux événements auxquels ils sont confrontés : leurs systèmes de valeurs, leurs repères normatifs, leurs interprétations de situations conflictuelles ou non, leurs lectures de leurs propres expériences* » (2011).

L'entretien consiste à recueillir de l'information de vive voix auprès des enquêtés qui témoignent de leurs attitudes, leurs comportements, leurs représentations ou des faits qu'ils ont observés. Ce mode de collecte de données se prépare avent sa mise en œuvre et peut être réalisé en personne ou par téléphone. Le chercheur prépare son entretien en formulant des questions qui vont l'orienter dans sa collecte et en sélectionnant des enquêtés qui possèdent des informations sur son objet de

recherche. A. Anger indique que « *le chercheur interroge telle personne parce que cette personne possède telle caractéristique, parce qu'elle appartient à telle couche sociale, parce qu'elle a connue tel type d'expérience* » (1997).

Dans un premier temps, le chercheur doit procéder préalablement à la formulation des questions de son entretien avant sa mise en place, en le centrant autour de ses hypothèses. Ensuite, il sélectionne des personnes qui estiment être en possession d'informations sur son objet de recherche. En d'autres termes « *le chercheur interroge telle personne parce que cette personne possède telle caractéristique, parce qu'elle appartient à telle couche sociale, parce qu'elle a connue tel type d'expérience* » (A. Anger, 1997 :15). De manière générale, il existe quatre types d'entretiens : *L'interview centrée, l'interview non directive, l'interview semi-directive, l'interview directive.*

2. Types d'entretien

Avant de mettre en œuvre un entretien dans une recherche, il est essentiel de déterminer le type d'entretien le plus adapté aux objectifs poursuivis et à la nature des données recherchées. En effet, comme le soulignent les travaux en méthodologie, il existe plusieurs formes d'entretiens qui se distinguent par le degré de liberté accordé à l'enquêté et par le niveau de structuration imposé par le chercheur .

Le choix entre ces différents types dépend du cadre théorique de la recherche, des hypothèses formulées et du type d'informations que l'on souhaite recueillir (exploratoires, explicatives ou comparatives). Certains entretiens privilégient la spontanéité et l'expression libre des enquêtés, tandis que d'autres reposent sur un guidage plus strict afin d'obtenir des données précises et comparables. Ainsi, on distingue généralement plusieurs types d'entretiens, allant des formes les plus ouvertes aux plus structurées, chacun présentant des avantages et des limites en fonction des objectifs de l'enquête.

2.1.L'entretien centré (ou exploratoire) : consiste à explorer en profondeur un thème préalablement défini en s'appuyant sur une liste de points à aborder (il n'y a pas de questions préétablies). Il permet de voir ce que les enquêtés ont à exprimer à propos d'une expérience ou d'un évènement vécu. Il peut se dérouler sur plusieurs phases.

2.2.L'entretien non directif : Dans ce type d'entretien, l'enquêteur annonce le thème principal

décomposé en sous-thèmes déterminés préalablement et à propos desquels les enquêtés parlent, sans que des questions directes ne leur soient posées. L'enquêté est libre d'organiser son discours comme il le souhaite et c'est au chercheur d'adopter une position neutre. Dans ce type d'entretien, le sujet sur lequel l'informateur va parler n'est pas délimité. De ce fait, l'informateur aborde le thème de l'enquête de façon générale (Blanchet et Gotman, 2010).

2.3.L'entretien semi-directif : il se place entre l'entretien directif et non directif dans la mesure où il accorde à l'enquêté assez de temps pour exprimer son point de vue librement et sans qu'il soit interrompu par l'enquêteur.

Le dernier recentrera simplement l'entretien par rapport à ses hypothèses et encouragera l'enquêté à donner plus d'informations sur le sujet traité. L'enquêteur doit recentrer ses questions afin de ne pas perdre de vue l'objectif qu'il s'est fixé (Blanchet et Gotman, 2010).

2.4.L'entretien directif : il ressemble à un questionnaire oral où l'espace de liberté accordé à l'enquêté est le plus réduit et les questions sont préétablies et non majoritairement improvisées comme les cas précédents. L'objectif de ce type d'entretien est de vérifier avec précision des points précis ou de recueillir les informations avec détail. C'est pour cette raison que l'enquêteur pose les mêmes questions à tous les interviewés. Cependant, à cause des limites que lui pose l'enquêteur, l'enquêté n'aura pas une grande liberté pour s'exprimer (Blanchet et Gotman, 2010).

En somme, la visée principale de l'entretien est de permettre de comprendre la signification de certaines pratiques communautaires, les points de vue des participants aux évènements et de reconstituer des événements passés ou des expériences vécues.

En effet, l'entretien est une sorte de discussion orientée qui présente de la souplesse pour sa confection peu directive tout en assurant une profondeur et une finesse de l'information désirée. Cependant, cette technique ne permet pas de prévoir le mode d'analyse des éléments recueillis et un biais peut également être introduit par l'intervieweur dans sa manière de poser les questions ou d'interpréter les réponses.

CONCLUSION

En somme, l'entretien constitue une méthode de collecte de données incontournable dans les approches qualitatives, en ce qu'il permet d'accéder à la complexité des pratiques, des discours et des représentations des acteurs. Sa souplesse et sa profondeur en font un outil particulièrement

adapté à l'analyse des phénomènes linguistiques contextualisés. Toutefois, son efficacité repose sur une préparation rigoureuse, une posture d'écoute attentive et une maîtrise des techniques d'enquête. Malgré certaines limites, notamment liées aux biais possibles de l'enquêteur ou à la difficulté d'analyse des données recueillies, l'entretien demeure un dispositif méthodologique fondamental. Bien conduit, il permet non seulement de vérifier des hypothèses, mais aussi de faire émerger de nouvelles pistes de réflexion, contribuant ainsi à enrichir la compréhension des réalités étudiées.

APPLICATION : -Elaborez une fiche technique sur la conduite d'un entretien dans un travail de recherche en sociolinguistique.

COURS 11 : OUTILS ET TECHNIQUES DE COLLECTE DES DONNEES :« L'ANALYSE DE CONTENU »

INTRODUCTION

En sciences sociales, l'analyse du discours s'est développée dans un rapport critique avec l'analyse de contenu, « une technique de recherche pour la description objective, systématique et quantitative du contenu manifeste de la communication » (Bardin 1993 : 21 cité dans Charaudeau et Maingueneau 2002 : 39) reposant sur une catégorisation thématique des extraits textuels et leur traitement statistique. Conçue comme une méthodologie distincte, l'analyse du discours s'appuie sur des marques linguistiques pour établir son interprétation du sens. En outre, elle procède à une analyse des marques énonciatives attestant des actes de langage et de l'ancrage contextuel du discours.

1.L'ANALYSE DE CONTENU

L'analyse de contenu est une technique qui permet d'effectuer des études détaillée sur des informations, des documents et des témoignages qui sont manifestement complexes. Pour Pinto et Grawitz, 1969,

> « *l'analyse de contenu est une technique de recherche pour la description objective, systématique et quantitative du contenu manifeste des communications ayant pour but de les interpréter* ».

Il s'agit donc d'une technique *objective* (poursuivre les mêmes objectifs pour aboutir aux mêmes résultats), *systématique (*un contenu recensé répertorié et exploré) *et quantitative* (s'appuie sur des calculs, comparaisons et statistiques).

Cette technique couvre un champ très étendu d'application et tend de dégager des significations, associations et intentions non directement décelables à la simple lecture des données.

De ce fait, elle peut servir à traiter des thématiques portant sur :

- Les représentations sociales et les pratiques discursives d'une communauté,
- La réalisation graphique, l'enchaînement discursif, l'expression et l'énonciation,
- L'étude structurale des interactions recueillies.

En somme, cette technique met en ouvre des procédures analytiques relativement précises qui permettent aux chercheurs d'étudier tout document de nature communicationnelle (journaux, rapports, émissions de radio ou de télévision, discours, etc.) ou produit par le chercheur lui-même

(questionnaire, tests verbaux, interviews, etc).

Les types d'analyse de contenu : il existe six types d'analyse de contenu

- **L'analyse d'exploration :** consiste à explorer un terrain ou une communauté afin de déterminer des hypothèses et des orientations.
- **L'analyse de vérification :** consiste à vérifier la justesse et la finesse des hypothèses préétablies, en s'appuyant sur des objectifs bien déterminés.
- **L'analyse qualitative :** consiste à rechercher des mots, des concepts des notions dans un corpus précis afin de comprendre l'effet et l'impact sémantique de ces éléments dans un contexte donné (leur fréquence ne contribue en rien dans ce type d'analyse.

*« Ma recherche s'inscrit plus particulièrement dans une démarche d'analyse de contenu. L'analyse de contenu a comme objectif de « connaître la vie sociale .Elle est une méthode qui procède à partir de documents écrits afin d'observer la pensée des sujets sa dimension sociale. En d'autres mots, cette méthode étudie la production textuelle (discours oraux ou écrits) en vue d'en faire une analyse interprétative des documents. Elle envisage le texte et la lectur*e comme « le lieu privilégié d'observation de l'élaboration du sens social » (ibid : 360).

L'analyse de documents textuels est un mode d'accès à l'étude de la vie sociale et mène à la construction d'un discours savant d'interprétation des documents à partir du langage des sujets et des interprétations du monde de ceux-ci (ibid: 358 »

Conclusion

Chaque méthode a ses propres forces et ses limites, et le choix de la méthode dépendra de la question de recherche, des ressources disponibles et des objectifs de l'étude. Il est souvent recommandé de combiner deux méthodes pour trianguler les données et obtenir une compréhension plus complète du phénomène étudié.

Tableau Récapitulatif

Méthode	**Objectif principal**	**Type de données**	**Avantages**	**Inconvénients**
Analyse de contenu	Décrire le contenu de communications	Quantitatives et qualitatives	Analyse de grandes quantités de données, identification de tendances	Peut être subjective, dépend de la qualité des documents
Observation	Observer les comportements et les interactions	Qualitatives et quantitatives	Données directes sur les comportements, compréhension du contexte	Peut être intrusive, biais de l'observateur
Questionnaire	Recueillir des données standardisées auprès d'un grand nombre de personnes	Quantitatives et qualitatives	Données standardisées, facile à administrer	Manque de flexibilité, taux de réponse variable

COURS12: OUTILS, TECHNIQUES DE COLLECTE DE DONNEES EN LINGUISTIQUE : L'expérimentation

INTRODUCTION

Dans le champ des sciences du langage, la collecte des données constitue une étape déterminante dans la construction et le traitement du corpus. Parmi les différentes méthodes mobilisées (observation, enquête, analyse de corpus), l'expérimentation se distingue par sa rigueur méthodologique et sa capacité à établir des relations causales entre les phénomènes étudiés. Contrairement aux approches descriptives qui se limitent à observer les faits tels qu'ils se présentent, l'expérimentation repose sur une intervention contrôlée du chercheur, visant à tester des hypothèses précises.

Ainsi, l'expérimentation s'inscrit dans une démarche hypothético-déductive héritée des sciences expérimentales, notamment des travaux fondateurs de Claude Bernard (1856), et adaptée aux spécificités des sciences humaines. Elle permet, dans le domaine linguistique, d'analyser les comportements langagiers en fonction de variables manipulées, et de mieux comprendre les mécanismes sous-jacents à la production et à la compréhension du langage.

1. Définition et fondements de la méthode expérimentale

L'expérimentation est généralement considérée comme l'une des techniques les plus efficaces en recherche, dans la mesure où elle permet de mettre en évidence des relations de cause à effet entre différentes variables. Selon M. Grawitz (2011), elle peut être définie comme une « observation provoquée » portant sur une situation construite et contrôlée par le chercheur, dans le but de vérifier ou d'infirmer une hypothèse issue d'un cadre théorique préalable.

Dans cette perspective, l'expérimentation ne se réduit pas à la simple production d'une situation artificielle ; elle constitue avant tout une mise à l'épreuve systématique d'une hypothèse. Elle implique donc l'existence d'un socle théorique initial, à partir duquel le chercheur formule des hypothèses qu'il soumet ensuite à vérification empirique.

Elle est qualifiée comme la plus efficace des techniques, dans la mesure où elle permet la mise en évidence des relations causales existant entre diverses variables (montrer à quel point la manipulation d'une variable peut entraîner des effets au niveau d'une autre variable) afin de parvenir

à extraire les lois générales. Pour M. Grawitz (2011), l'expérimentation est une observation provoquée portant sur une situation créée et contrôlée par le chercheur, et qui a pour but de valider (ou invalider) une ou des hypothèses issues d'un système théorique.

En effet, l'expérimentation implique l'application de la démarche définie par Claude Bernard (1856), et qui exige la possession d'un corpus théorique de départ bien déterminé (collecter des faits), qui sera diffèrent d'une enquête pour parvenir à tester une hypothèse causale concernant la mesure d'un phénomène précis en fonction de la manipulation (la variation) d'un ou plusieurs facteurs.

L'expérimentation n'est donc pas seulement la production d'une expérience ; c'est encore et principalement la mise à l'épreuve d'une hypothèse par une expérience.

Cette méthode consiste à insérer une variable déjà étudiée, tout en contrôlant les autres. Les résultats de cette méthode expérimentale sont évalués par rapport à un groupe témoin. C'est-à-dire que les chercheurs ne valident les conclusions que par à une expérience strictement identique.

Se fondant sur l'observation, l'objectif de la méthode expérimentale est la comparaison entre groupes équivalents de participants en manipulant une ou plusieurs variable(s) provoquant ainsi une variation des réponses des participants , ce qui va permettre d'expliquer les différences observées entre les groupes .

En effet, il l existe deux types de variables : *la variable indépendante* est une caractéristique portant sur l'individu, l'environnement et qui est manipulée par le chercheur afin de contrôler ou d'analyser son impact sur le comportement cible.

La variable dépendante : est une caractéristique qui dépend du changement des variables indépendantes et permet de tester son impact en comparant les résultats des groupes étudiés.

Le chercheur peut procéder à un contrôle *expérimental* ou *statistique*. Il s'agit, selon G. Borst et A. Gachia (2016) de manipuler une variable indépendante et de mesurer l'effet produit par cette manipulation sur une variable dépendante, tout en contrôlant les variables externes qui pourraient potentiellement avoir une influence. L'effet est mesuré, dans un premier temps, à l'aide de statistiques descriptives. La démarche expérimentale au niveau méthodologique peut être caractérisée par deux étapes

2. Spécificité de la démarche expérimentale en linguistique

Contrairement à l'enquête ou à l'observation naturaliste, l'expérimentation suppose une intervention active du chercheur sur les conditions de production des données. Elle consiste à manipuler une ou plusieurs variables afin d'en observer les effets sur un phénomène linguistique donné.

Cette méthode repose sur la constitution de situations expérimentales contrôlées, souvent comparées à des situations témoins. L'objectif est de garantir que les différences observées dans les résultats sont bien dues à la variable manipulée, et non à des facteurs externes. Ainsi, les conclusions ne sont validées que si l'expérience peut être reproduite dans des conditions strictement identiques.

3. Les variables dans l'expérimentation

Le cœur de la méthode expérimentale réside dans la manipulation et le contrôle des variables. On distingue principalement deux types de variables :

- **La variable indépendante** : il s'agit du facteur manipulé par le chercheur. Elle peut concerner une caractéristique de l'individu (âge, niveau linguistique), de l'environnement (contexte de communication), ou du stimulus (type de texte, structure syntaxique, etc.). Son rôle est de produire un effet observable.
- **La variable dépendante** : elle correspond au phénomène mesuré, c'est-à-dire à la réponse des participants à la manipulation expérimentale. Elle dépend directement des variations de la variable indépendante et permet d'évaluer l'effet de celle-ci.

Le chercheur doit également prendre en compte les variables dites « parasites » ou externes, susceptibles d'influencer les résultats. Leur contrôle est essentiel pour garantir la validité de l'expérimentation.

4. Le contrôle expérimental et l'analyse des résultats

L'un des principes fondamentaux de l'expérimentation est le contrôle des conditions dans lesquelles se déroule l'étude. Comme le soulignent G. Borst et A. Gachia (2016), il s'agit de manipuler une variable indépendante tout en mesurant ses effets sur une variable dépendante, en neutralisant autant que possible les influences extérieures.

Ce contrôle peut être :

- **expérimental**, par la mise en place de groupes équivalents (groupe expérimental vs groupe témoin) ;

- **statistique**, à travers des outils d'analyse permettant d'isoler les effets des variables pertinentes.

Les résultats obtenus sont d'abord traités à l'aide de statistiques descriptives (fréquences, moyennes, écarts), puis peuvent faire l'objet d'analyses plus approfondies afin de dégager des régularités significatives.

5. Les étapes de la démarche expérimentale

Sur le plan méthodologique, la démarche expérimentale peut être structurée en deux grandes étapes :

1. **La phase de conception** : elle consiste à formuler une hypothèse, définir les variables, construire le protocole expérimental et sélectionner les participants.
2. **La phase de réalisation et d'analyse** : elle implique la mise en œuvre de l'expérience, la collecte des données, leur traitement et l'interprétation des résultats.

CONCLUSION

Ces étapes garantissent la cohérence et la validité scientifique de la recherche.

En définitive, l'expérimentation constitue une méthode centrale en linguistique, particulièrement adaptée à l'étude des mécanismes cognitifs et des comportements langagiers. Elle se caractérise par la manipulation contrôlée de variables, la comparaison entre groupes, et l'analyse rigoureuse des effets observés.

Son principal atout réside dans sa capacité à établir des relations causales, ce qui en fait un outil privilégié pour tester des hypothèses et valider des modèles théoriques. Toutefois, elle exige une grande rigueur méthodologique, notamment dans le contrôle des variables et la reproductibilité des résultats.

Ainsi, bien qu'elle ne remplace pas les méthodes comme l'analyse de corpus ou l'enquête, l'expérimentation s'inscrit comme un complément essentiel, contribuant à une compréhension plus fine et plus scientifique des phénomènes linguistiques.

CHAPITRE III : TRAITEMENT DES DONNEES DE CORPUS : MTHODES ET APPROCHES D'ANALYSE EN LINGUISTIQUE

- CHAPITRE III : TRAITEMENT DES DONNÉES DE CORPUS : MÉTHODES ET APPROCHES D'ANALYSE EN LINGUISTIQUE

Objectifs du chapitre :

Ce chapitre vise à développer les compétences analytiques des étudiants dans le traitement et l'interprétation des données issues de corpus.

- Savoir structurer, annoter et organiser un corpus en vue de son analyse.
- Maîtriser les principes de l'analyse quantitative et qualitative des données linguistiques.
- Comprendre les étapes de l'interprétation scientifique des résultats issus de corpus.
- Identifier et mobiliser différentes approches théoriques d'analyse linguistique (énonciative, pragmatique, lexicométrique, dialogique, sémiologique, etc.).
- Savoir adapter une méthode d'analyse en fonction de la problématique de recherche.
- Développer une capacité d'interprétation critique des données linguistiques.
- Comprendre la complémentarité entre approches théoriques et analyses empiriques.

COURS 13 : TRAITEMENT DES DONNEES : STRUCTURER, ANNOTER ET ANALYSER UN CORPUS

Introduction

Dans toute recherche en linguistique de corpus, la collecte des données ne constitue qu'une étape préliminaire. La véritable valeur scientifique du corpus se construit lors de son traitement, c'est-à-dire au moment où les données brutes sont organisées, enrichies et analysées en fonction d'un cadre théorique précis. Cette phase, essentielle, engage le chercheur dans un travail d'interprétation rigoureux visant à transformer des occurrences linguistiques en résultats exploitables.

Ainsi, le traitement des données repose sur trois opérations fondamentales : la structuration du corpus, son annotation, et enfin son analyse. Ces étapes permettent non seulement de rendre les données lisibles et manipulables, mais aussi de dégager des régularités, des variations et des relations significatives entre les phénomènes linguistiques observés. Dans cette perspective, la synthèse interprétative apparaît comme l'aboutissement d'un processus analytique complexe, articulant description fine et compréhension globale.

1. De la collecte à la structuration du corpus

Une fois les données recueillies à partir de diverses techniques (enquêtes, observations, questionnaires, interactions enregistrées), le chercheur entre dans une phase de traitement qui consiste d'abord à organiser ces données. Cette structuration implique un tri, un classement et une mise en forme des informations recueillies afin de constituer un corpus cohérent et exploitable.

Chaque donnée est examinée avec précision, puis intégrée dans un ensemble descriptif où elle prend sens en relation avec d'autres éléments. Ce travail ne se limite pas à une simple organisation formelle : il suppose également une réflexion sur les unités d'analyse pertinentes (mots, phrases, tours de parole, actes de langage, etc.) en fonction des objectifs de recherche.

2. Annotation du corpus et enrichissement des données

L'annotation constitue une étape clé du traitement des données. Elle consiste à ajouter des informations linguistiques ou contextuelles aux données brutes, afin de faciliter leur analyse. Ces annotations peuvent être de différents types : morphologiques, syntaxiques, sémantiques, pragmatiques ou encore discursives.

Dans le cadre de la linguistique de corpus, l'annotation permet notamment de repérer des phénomènes récurrents, d'identifier des catégories linguistiques et de rendre possible l'exploitation automatique ou semi-automatique des données. Elle contribue ainsi à transformer un corpus brut en un corpus enrichi, apte à soutenir des analyses approfondies.

3. Approches analytiques : entre description structurale et contextualisation

Les procédures d'analyse en linguistique reposent sur des traditions méthodologiques bien établies. Les approches structurales (phonologique, morphologique, syntaxique) permettent d'identifier les régularités formelles du système linguistique, en mettant en évidence les oppositions et les variations significatives.

Cependant, dans une perspective contemporaine, notamment en linguistique de corpus et en analyse du discours, l'accent est mis sur la contextualisation des données. Il ne s'agit plus seulement d'analyser des unités isolées, mais de les inscrire dans des ensembles discursifs plus larges, en tenant compte des paramètres sociaux, culturels et interactionnels.

Ainsi, l'analyse porte autant sur ce qui est stable que sur ce qui varie, qu'il s'agisse de variations diachroniques, sociolinguistiques ou discursives. Cette approche intégrative permet d'articuler les dimensions formelles et fonctionnelles du langage.

4. L'analyse du discours et l'intégration du contexte

Dans le domaine de l'analyse du discours, les méthodes privilégient une approche globale, centrée sur les relations entre les énoncés, leur contexte de production et les pratiques sociales dans lesquelles ils

s'inscrivent. Les techniques de cooccurrence, d'analyse pragmatique et d'étude des positionnements énonciatifs permettent de saisir les stratégies discursives des locuteurs.

Par ailleurs, l'analyse ne se limite pas au verbal : elle intègre également les dimensions non verbales (gestes, mimiques, postures), particulièrement pertinentes dans les interactions orales. Ces éléments contribuent à la construction du sens et doivent être pris en compte dans une perspective multimodale. Les travaux en sociolinguistique interactionnelle et en ethnographie de la communication (notamment dans la lignée de Gumperz et Hymes) insistent sur la capacité des locuteurs à interpréter les indices contextuels et à mobiliser leurs compétences linguistiques et socioculturelles dans l'interaction.

5. Analyse qualitative des données de corpus

Dans cette méthodologie, l'analyse qualitative occupe une place centrale. Elle vise à comprendre les phénomènes linguistiques dans leur complexité, en tenant compte du contexte et des pratiques sociales. L'analyse quantitative, quant à elle, intervient comme un outil complémentaire, permettant de mesurer des fréquences, de repérer des tendances ou de valider certaines observations. Toutefois, les spécialistes reconnaissent que les données chiffrées ne prennent pleinement sens que lorsqu'elles sont interprétées à la lumière du contexte. Ainsi, l'analyse textuelle et discursive dépasse le simple comptage pour devenir une véritable démarche interprétative.

6.La synthèse interprétative

La synthèse interprétative constitue l'aboutissement du traitement des données. Elle ne se réduit pas à un résumé des résultats, mais correspond à une mise en relation des différents niveaux d'analyse dans une perspective systémique. Elle repose notamment sur :

- la confrontation des données objectives (observées dans le corpus) et des données subjectives (issues des locuteurs) ;
- la recherche de régularités et de phénomènes récurrents ;
- l'analyse des relations entre variables ;
- la comparaison entre résultats observés et hypothèses initiales.

L'interprétation intervient à deux niveaux : d'une part, dans la compréhension des mécanismes de production et d'interprétation des énoncés ; d'autre part, dans l'analyse des stratégies mobilisées par les locuteurs en contexte.

- **L'analyse quantitative** Dans le cadre d'une approche quantitative, le traitement des données s'organise autour de trois opérations principales :

a-Description et agrégation des données

Il s'agit d'identifier les caractéristiques des variables présentes dans les données et de les organiser de manière à faire apparaître leur distribution.

b. Analyse des relations entre variables

Cette étape vise à mettre en évidence les relations de dépendance ou d'indépendance entre les variables, en lien avec les hypothèses formulées.

c. Comparaison et interprétation des écarts

Le chercheur confronte les résultats obtenus aux hypothèses initiales, mesure les écarts et propose des interprétations, pouvant conduire à la reformulation des hypothèses.

Conclusion

Le traitement des données en linguistique de corpus constitue une étape déterminante dans la production du savoir scientifique. À travers la structuration, l'annotation et l'analyse des corpus, le chercheur transforme des données brutes en objets d'étude interprétables, en articulant rigueur méthodologique et réflexion théorique.

La synthèse interprétative, au cœur de cette démarche, permet de dépasser la simple description pour accéder à une compréhension globale et contextualisée des phénomènes linguistiques. En combinant approches qualitatives et quantitatives, et en intégrant les dimensions discursives, sociales et interactionnelles, la linguistique de corpus s'affirme ainsi comme un champ dynamique, capable de rendre compte de la complexité des pratiques langagières contemporaines.

COURS 14: LA SYNTHESE INTERPRETATIVE EN LINGUISTIQUE DE CORPUS : UNE APPROCHE SYSTEMIQUE

Introduction

Dans le processus de traitement des données en linguistique de corpus, l'analyse ne constitue pas une finalité en soi. Elle doit être prolongée par une étape essentielle : la synthèse interprétative. Celle-ci permet de dépasser la simple description des faits linguistiques pour accéder à une compréhension globale et contextualisée des phénomènes observés.

Contrairement à une vision réductrice qui assimilerait la synthèse à un simple résumé, la synthèse interprétative relève d'une démarche systémique, où les données sont mises en relation entre elles, mais aussi avec leur contexte de production, les hypothèses de recherche et les cadres théoriques mobilisés. Elle constitue ainsi un moment clé de la construction du sens scientifique.

1. Définition de la synthèse interprétative

Une fois le recueil des données effectué, en allant de la prise globale d'informations et d'indices de micro-situations de l'échantillon constitués par enquêtes directives, par observation participante, de questionnaire, arrive la phase la plus fine de la démarche, **l'analyse de ces données** ouvrant aussitôt sur l'élargissement que va produire la synthèse interprétative .

Cette étape analytique fonctionne de faits ou chaque donnée est examiné en détail, trié, rangé dans un ensemble descriptif dont telle forme est une pièce. Cependant, et c'est là notamment qu'une approche systémique complexe interprétative diffère d'une approche analytique structurale classique, le travail du chercheur ne s'arrête pas là.

Du point de vue méthodologique, en effet, en cohérence avec le cadre théorique, le fonctionnement global de l'ensemble agit sur les parties. C'est le contexte global de la ou des situations de communication

Les procédures d'analyse mises en point depuis de nombreuses années dans les champs linguistique, sociologiques, anthropologiques…etc sont connues et établies. **Il s'agit des procédures phonologiques, morphologiques, syntaxiques, etc**. Des méthodes d'échantillonnage et de calcul statistique d'investigation structurale des langues notamment en linguistique. En matière **d'analyse**

linguistique, on y recherche ce qui varie au moins autant que ce qui est stable, que ce soit dans une structure donnée (**il n'ya par exemple opposition phonologique que s'il y'a variation signifiante possible).** Ou dans les pratiques linguistiques en général (attention portée aux changements diachroniques ou aux diverses variations /variétés non standard d'une langue) , **une approche phonologique, combinée à une approche éthno-sociolinguistique.**

La synthèse interprétative correspond donc à une mise en cohérence des résultats analytiques dans une perspective globale. Elle vise à répondre à la problématique de recherche en articulant différents niveaux d'analyse :

- linguistique (formes, structures),
- discursif (organisation des énoncés),
- sociolinguistique (variations, usages),
- contextuel (situations de communication).

2.Etapes et enjeux de la synthèse interprétative

Elle suppose:

A- **une lecture transversale du corpus**, où le chercheur ne se limite plus à isoler des faits, mais cherche à comprendre leurs interactions et leurs significations.

B- Une **Articulation entre données objectives et subjectives**

Un des fondements de la synthèse interprétative réside dans la confrontation entre :

- **données objectives** : occurrences linguistiques observées dans le corpus (formes, fréquences, structures),
- **données subjectives** : représentations, intentions ou perceptions des locuteurs (issues d'entretiens, questionnaires, commentaires).

c- une recherche de régularités et de phénomènes récurrents : La synthèse implique l'identification de **schémas récurrents** dans les données :

répétition de structures, stratégies discursives fréquentes, des formes linguistiques dominantes.

Exemple (réseaux sociaux) Dans un corpus de commentaires YouTube :

- usage récurrent de l'anglais dans des phrases françaises (« c'est amazing », « trop cringe »).

d- Analyse des relations entre variables

La synthèse interprétative examine les **relations entre différents facteurs** :

- variables linguistiques,
- variables sociales (âge, niveau, contexte),
- variables discursives.

e- Confrontation aux hypothèses et interprétation des écarts

La synthèse implique une comparaison entre :

- résultats observés
- hypothèses initiales

Les écarts sont particulièrement riches, car ils ouvrent de nouvelles perspectives et pistes pour de nouvelles recherches.

3. Les deux niveaux de l'interprétation

L'interprétation fonctionne à deux niveaux complémentaires :

a. Niveau micro (énonciatif)

Compréhension des mécanismes de production du sens :

- choix lexicaux,
- structures syntaxiques,
- indices pragmatiques.

b. Niveau macro (contextuel)

Analyse des stratégies des locuteurs :

- positionnement social,
- intentions communicatives,
- inscription dans un contexte culturel.

Exemple (réseaux sociaux)

Un commentaire ironique : Niveau micro : marqueurs linguistiques d'ironie

Niveau macro : critique implicite d'un contenu ou d'une idéologie

En termes d'analyse de discours, on préfère bien sur les techniques de contextualisation (rapport à l'ensemble d'un discours et à d'autres discours, ainsi qu'aux paramètres sociaux » aux

décompositions sémantiques et structuralistes et aux analyses limitées à la phrase ou au mot.

Une analyse quantitative à l'aide de chiffres joue un rôle mineur dans l'ensemble de cette méthodologie essentiellement qualitative, ou ils ne servent que de compléments d'enquêtes. Les spécialistes des analyses quantitatives conviennent eux-mêmes du fait que « au-delà du vocabulaire, la prise en compte du contexte peut dans la mesure ou c'est possible changer complètement la perspective et permettre une analyse de toute une autre ampleur. L'analyse textuelle devient alors en elle-même une blanche de l'analyse des données » .

Les procédures comparatives contribuent à réduire l'implication subjective du chercheur et offre des résultats satisfaisants en termes de compréhension d'un cas précis qu'en terme de généralisation c'est-à-dire la synthèse, conçue comme « *la saisie globale des paramètres en jeu dans une perspective systémique complexe. Il ne s'agit pas d'un résumé final ,mais des procédures analytiques comparatives et interprétatives* ». L'interprétation, dans notre domaine, fonctionne à deux niveaux. Au premier niveau, elle désigne l'objet de la sociolinguistique interactionnelle et ethnographie de la communication de Gumperz et hymes « *il s'agir de découvrir les liens étudiés entre la perception des signes linguistiques de surface et l'interprétation d'expliquer la capacité humaine à contextualiser l'interprétation*…de mettre l'accent sur les stratégies qui gouvernent l'emploi par l'acteur des connaissances lexicales, grammaticales, sociolinguistiques et autres dans la production et l'interprétation des messages en contexte ».

Ce sont les interprétations par les interlocuteurs des signaux ethno-sociolinguistiques (variation) comme indices signifiants dans une interaction langagière. Cette synthèse interprétative se réalise en :

-mise en relation de données langagières observées « objective » avec le contexte perçu de leur productions et avec d'autre part les données « subjectives » recueillies auprès des locuteurs.

-Recherche de récursivité des phénomènes et des fonctionnements. Analyse des relations entre variables

- Comparaison des résultats observés avec ceux attendus et interprétation des écarts : Comparaison des parties et du tout, selon données jusqu' à structuration. confronter les relations avec celles prévues dans les hypothèses.

Synthèse

En effet, dans le cadre d'une recherche quantitative le chercheur recourt aux trois opérations de l'analyse des informations et qui se présentent ainsi :

a- **Décrire les données et les agréger (si nécessaire) :** cette opération consiste à mettre en avant, lors de la description les caractéristiques des variables présentes dans les données à analyser , et à les présenter d'une manière (agrégée si nécessaire) qui s'apparente à la distribution des

variables indiquées dans les hypothèses (procéder à une répartition en sous-groupe ou les décrire en fonction de nouvelles données pertinentes). Au terme de cette phase le chercheur aura un inventaire de données manifestant des variables bien mises en exergue.

b- **Analyse des relations entres les variables :** cette étape se donne pour objectif de mettre en lumière, en s'appuyant sur les hypothèses, les relations pouvant exister entre les variables des données observées. De ce fait, au cours de cette de phase, le chercheur est appelé à manifester avec rigueur, et comme prévu dans les hypothèses, toute relation de dépendance, d'indépendance ou de cohérence observée entre les variables. Et pour marquer la rigueur des hypothèses proposées, il faut passer par les hypothèses complémentaires qui confirment la fiabilité des relations prévues par les hypothèses.

c- **Comparaison des résultats observés avec ceux attendus et interprétation des écarts :** la visée principale de cette étape est de confronter les relations déterminées lors de l'observation avec celles prévues dans les hypothèses. Il s'agit ainsi de mesurer les écarts entre ces relations étudiées et de les interpréter dans un second temps. Au terme de cette étape, le chercheur devra déterminer avec précision la nature de ces relations afin de pouvoir confirmer ou infirmer ses hypothèses.

En revanche, il est à souligner que, cas échéant, le chercheur est appelé à présenter des explications en ce qui concernes les écarts qui ne riment pas avec les hypothèses prévues, ou à prendre l'initiative d'élaborer de nouvelles hypothèses pour en tirer des conclusions pertinentes.

CONCLUSION

La synthèse interprétative constitue une étape décisive en linguistique de corpus, en ce qu'elle permet de transformer des résultats analytiques en connaissances significatives. En articulant données, contexte et hypothèses, elle offre une vision globale et dynamique des phénomènes langagiers.

Elle révèle également la complexité du langage, en montrant que les faits linguistiques ne peuvent être compris indépendamment des pratiques sociales et des contextes d'usage. Ainsi, elle s'impose comme un outil essentiel pour toute recherche visant à saisir la dimension vivante et située du langage.

COURS 15 : L'ANALYSE QUANTITATIVE DES CORPUS LINGUISTIQUES

INTRODUCTION

Si la linguistique de corpus est souvent associée à une approche qualitative, elle intègre également des outils quantitatifs indispensables à l'analyse des données. L'analyse quantitative permet de mesurer, comparer et objectiver les phénomènes linguistiques, en apportant une dimension empirique et vérifiable aux observations.

Toutefois, loin de se suffire à elle-même, elle doit être articulée à une interprétation contextualisée. Ce cours propose d'explorer les principales étapes de l'analyse quantitative et leur application dans des contextes concrets.

1. Description des données linguistiques

Cette première étape consiste à :

- identifier les variables présentes dans le corpus,
- organiser les données,
- produire des distributions (fréquences, proportions).

Exemple. Pour un c**orpus de copies d'étudiants :**

- **Variable** : temps verbaux utilisés permet une première cartographie des usages
- 60 % présent
- 30 % passé composé
- 10 % imparfait

2. Analyse des relations entre variables : Cette étape vise à comprendre les liens entre différentes variables :

- dépendance (une variable influence l'autre),
- indépendance,
- corrélation.

Exemple (classe)

Variables :niveau des apprenants /type d'erreurs

débutants → erreurs morphologiques

avancés → erreurs discursives

Donc, une évolution des difficultés selon le niveau.

3. Comparaison et interprétation des écarts

Le chercheur compare :

- résultats observés,
- hypothèses initiales.

Les écarts sont analysés pour affiner la recherche.

Exemple (réseaux sociaux)

Hypothèse : forte dominance du français

Résultat : présence importante de l'arabe dialectal et de l'anglais

pratiques plurilingues,

identité hybride des locuteurs.

4. Limites de l'analyse quantitative

L'analyse quantitative présente certaines limites :

- réduction de la complexité,
- absence de contexte,
- risque de surinterprétation des chiffres.

Nous devons l'articuler avec une analyse qualitative.

5. Vers une approche mixte (quantitative et qualitative)

Les recherches contemporaines en linguistique de corpus tendent à dépasser l'opposition classique entre approches quantitative et qualitative, en privilégiant une démarche intégrative qui articule ces deux perspectives. En effet, chacune présente des apports spécifiques mais également des limites : l'analyse quantitative permet d'identifier des régularités, de mesurer des fréquences et de dégager des tendances générales à partir de grands ensembles de données, tandis que l'analyse qualitative offre les outils nécessaires pour interpréter ces résultats, en les replaçant dans leur contexte discursif, social et interactionnel. Ainsi, loin d'être concurrentes, ces deux approches apparaissent comme complémentaires et indissociables dans une démarche scientifique rigoureuse.

Dans cette perspective, l'analyse quantitative constitue souvent une première étape exploratoire. Elle permet de repérer des phénomènes saillants dans le corpus, tels que la fréquence d'un mot, d'une structure ou d'un procédé linguistique donné. Ces résultats chiffrés orientent ensuite l'analyse vers des zones pertinentes du corpus, qui feront l'objet d'une investigation qualitative plus approfondie. Cette

seconde étape vise à comprendre les fonctions, les valeurs et les enjeux des phénomènes observés, en tenant compte des paramètres contextuels et des intentions des locuteurs.

Un exemple particulièrement éclairant de cette complémentarité peut être observé dans l'étude des pratiques langagières en contexte numérique, notamment dans des corpus combinant productions d'apprenants de FLE et interactions sur les réseaux sociaux. Une analyse quantitative peut révéler une fréquence importante d'emprunts à l'anglais dans les productions écrites ou les commentaires en ligne (par exemple : *challenge*, *feedback*, *story*, *like*). Toutefois, ces données brutes ne suffisent pas à en expliquer la signification.

C'est à ce niveau qu'intervient l'analyse qualitative, qui permet d'interpréter ces emprunts en tant que marqueurs discursifs et identitaires. En effet, leur usage peut traduire une volonté d'inscription dans une culture numérique globalisée, une stratégie de valorisation sociale, ou encore une forme d'appropriation linguistique liée aux pratiques plurilingues des locuteurs. Ainsi, ce qui apparaît d'abord comme une simple donnée statistique se révèle porteur de significations complexes lorsqu'il est analysé dans son contexte d'usage.

En définitive, l'approche mixte permet de concilier la rigueur des données chiffrées et la richesse de l'interprétation. Elle favorise une compréhension plus fine et plus complète des phénomènes linguistiques, en articulant description et explication. Cette complémentarité méthodologique s'impose aujourd'hui comme une exigence dans les recherches en linguistique de corpus, notamment dans l'étude des pratiques langagières contemporaines .

CONCLUSION

L'analyse quantitative constitue un outil méthodologique essentiel en linguistique de corpus, permettant de structurer et d'objectiver les observations. Toutefois, elle ne prend tout son sens que lorsqu'elle est intégrée dans une démarche interprétative plus large.

En combinant mesures statistiques et analyse qualitative, le chercheur peut accéder à une compréhension fine et nuancée des pratiques langagières. Cette complémentarité fait de la linguistique de corpus une approche particulièrement adaptée à l'étude des usages contemporains du langage, notamment dans des contextes plurilingues et numériques.

COURS 16 : PANORAMA DES APPROCHES LINGUISTIQUES : L'APPROCHE COMMUNICATIONNELLE (MODELE DE JAKOBSON)

INTRODUCTION

Au-delà des approches énonciative et pragmatique qui ont marqué la linguistique et l'analyse de discours en particulier,. Et qui témoignent de la richesse et de la complexité inhérente à l'étude du langage. L'approche communicationnelle offre une perspective unique à travers laquelle les chercheurs explorent les dimensions variées du discours, mettant en lumière les structures qui composent la communication linguistique. Dans ce qui suit, nous aborderons ses principes.

1. L'approche communicationnelle : un modèle de Jakobson

Inspirée des travaux de Roman Jakobson dans les années 1960, elle propose une lecture du langage centrée sur ses fonctions dans l'acte de communication. Cette perspective marque un tournant important en linguistique, en mettant l'accent non seulement sur la structure du message, mais aussi sur les conditions de sa production et de sa réception.

L'intérêt de cette approche réside dans sa capacité à articuler les dimensions linguistiques, sociales et interactionnelles du discours. Elle permet ainsi de comprendre comment les éléments du langage s'organisent en fonction d'intentions communicatives spécifiques.

2. Composantes et fonctions du langage

Ce modèle considère les différentes fonctions du langage en association avec les composantes d'un acte de communication. Essentiellement axée sur l'importance du contexte et de la situation dans lesquels un discours émerge, cette approche cherche à décrypter les interactions sociales et à comprendre les composantes linguistiques ainsi que leurs relations dans des contextes diversifiés. Jakobson souligne la nécessité d'étudier le langage dans toutes ses fonctions pour saisir le fonctionnement d'une communication verbale. Chaque communication possède une fonction dominante et essentielle, distinguant ainsi un message d'un autre. Cependant, ces fonctions ne sont pas

mutuellement exclusives, elles se combinent et se superposent souvent. Il est rare qu'un message ait une seule de ces fonctions.

La structure verbale et les caractéristiques linguistiques d'un message sont profondément déterminées par sa fonction dominante, soulignant ainsi l'intention particulière qui guide la communication.

Destinateur et Fonction Expressive : Le destinateur, ou émetteur du message, utilise la fonction expressive pour informer le destinataire des aspects personnels, des intentions, et de la position de l'auteur par rapport au contenu.

- Message et Fonction Poétique : La fonction poétique donne au message une valeur esthétique propre, dépendant de l'art d'utiliser le langage pour exprimer des pensées et représenter le monde.

Exemple : Un poème utilise la fonction poétique pour créer une signification émotionnelle distincte en jouant avec les mots et les images, suscitant des émotions chez le lecteur.- Destinataire et la fonction conative: Le destinataire réagit en fonction de son interprétation de l'intention ou de la finalité communicative du destinateur. Exemple : Dans un discours commercial, le destinataire réagit aux arguments persuasifs du destinateur (la marque), influençant ainsi ses choix d'achat. 14 - Contexte et Fonction Référentielle : La fonction référentielle concerne le contexte et/ou le référent extralinguistique sur lequel repose le message. Exemple : Un article scientifique utilise la fonction référentielle pour décrire des faits, des expériences et des données, établissant un contexte objectif et informatif.- Contact et fonction phatique : La fonction phatique sert à établir, maintenir ou interrompre le contact physique et/ou psychologique entre l'émetteur et le récepteur du message. Exemple : Les salutations courantes dans une conversation (comme "Bonjour") remplissent une fonction phatique en établissant un contact social et en facilitant la communication.- Métalangage et Fonction Métalinguistique : La fonction métalinguistique se manifeste lorsque les acteurs utilisent la langue pour parler d'elle-même ou d'une autre langue. Exemple : Un enseignant expliquant la signification d'un terme à un élève utilise la fonction métalinguistique pour clarifier le sens du mot. L'analyse de discours selon l'approche communicationnelle ou fonctionnelle, inspirée par le modèle de Roman Jakobson, implique plusieurs étapes pour décortiquer les différentes composantes linguistiques et contextuelles d'un discours. Voici les étapes clés à suivre :- - - - Identification de la Fonction Dominante : Déterminez la fonction principale du discours en examinant comment les éléments linguistiques sont utilisés pour atteindre un objectif particulier. Les fonctions du langage selon

Jakobson comprennent le message, le contexte, le code, l'émetteur, le destinataire et la métalinguistique.

Analyse du Contexte : Étudiez le contexte dans lequel le discours a été produit. Cela englobe le cadre social, culturel et situationnel qui influe sur la communication. Comprenez comment ces facteurs contextuels affectent la signification du discours. Identification des Acteurs : Identifiez les émetteurs et les destinataires du discours. Analysez leurs rôles, leurs relations et leur influence potentielle sur le message. Examen du code Linguistique : Analysez les choix linguistiques, y compris le vocabulaire, la syntaxe, la grammaire et les figures de style. Comprenez comment ces éléments contribuent à la fonction dominante du discours.

Le modèle de Jakobson repose sur six éléments constitutifs de toute communication :

- le destinateur (émetteur),
- le destinataire (récepteur),
- le message,
- le contexte,
- le code,
- le contact.

À chacun de ces éléments correspond une fonction du langage récapitulées ci-dessous :

a- Fonction expressive (destinateur)

Elle permet à l'émetteur d'exprimer ses émotions, ses intentions et sa subjectivité. *Exemple : un discours politique où le locuteur exprime ses convictions.*

b- Fonction conative (destinataire) Elle vise à influencer ou à faire agir le destinataire. *Exemple : un message publicitaire incitant à l'achat.*

c- Fonction référentielle (contexte) Elle concerne la transmission d'informations objectives. *Exemple : un article scientifique décrivant des faits.*

d- Fonction phatique (contact) Elle sert à établir ou maintenir la communication. *Exemple : « Bonjour », « Tu m'entends ? »*

e- Fonction métalinguistique (code) Elle porte sur la langue elle-même. *Exemple : un enseignant expliquant un mot.*

f- Fonction poétique (message) Elle met l'accent sur la forme du message. *Exemple : un poème jouant sur les sonorités.*

3. Exemples de l'analyse communicationnelle des corpus politiques

Imaginons un débat politique où un candidat utilise la fonction expressive pour partager ses convictions et positions politiques. En utilisant un langage émotionnel et persuasif, le destinateur cherche à influencer positivement le destinataire (l'électorat). Le contexte politique et social sert de toile de fond (fonction référentielle), et les échanges verbaux créent un contact psychologique entre le candidat et les électeurs (fonction phatique). Ainsi, cette analyse permet de comprendre comment les différentes fonctions du langage interagissent pour façonner le discours politique dans un contexte donné.

L'analyse du discours selon cette approche repose sur plusieurs étapes :

- identifier la fonction dominante ;
- analyser le contexte de production ;
- repérer les acteurs et leurs rôles ;
- étudier les choix linguistiques.

Dans le discours politique, un candidat :

a) mobilise la fonction expressive (convictions),
b) utilise la fonction conative (persuasion),
c) s'inscrit dans un contexte socio-politique (fonction référentielle),
d) établit un lien avec le public (fonction phatique).

Cette combinaison montre comment les fonctions interagissent pour produire du sens.

CONCLUSION

En suivant ces étapes, l'analyse communicationnelle de discours vise à démystifier les mécanismes complexes qui sous-tendent la communication verbale, en mettant en lumière les choix linguistiques et contextuels qui façonnent le sens.

Elle offre ainsi un cadre d'analyse structuré permettant de décrypter les mécanismes fondamentaux de la communication verbale. En mettant en relation fonctions, acteurs et contexte, elle permet de comprendre comment le langage s'adapte aux intentions communicatives. Elle constitue ainsi un outil essentiel pour l'analyse de discours, notamment dans les domaines politique, médiatique et pédagogique.

COURS 17: L'APPROCHE ENONCIATIVE : CONCEPTS DE BASE

INTRODUCTION

L'émergence de l'approche énonciative en linguistique s'inscrit dans un tournant majeur opéré à partir des années 1960, marqué par le passage d'une linguistique centrée sur la langue comme système abstrait à une linguistique du discours, attentive aux conditions réelles d'usage du langage. Ce déplacement de perspective a conduit les chercheurs à considérer que le sens ne réside pas uniquement dans les structures linguistiques, mais qu'il se construit également dans et par les situations d'énonciation.

Dans le cadre de la linguistique de corpus, cette orientation s'avère particulièrement pertinente. En effet, l'analyse des unités linguistiques ne peut se faire indépendamment des facteurs extralinguistiques, tels que le locuteur, le contexte, le moment et l'intention communicative. L'approche énonciative permet précisément d'articuler ces dimensions en mettant en relation le discours et ses conditions de production.

1. Définition et fondements de l'approche énonciative

L'intérêt porté à l'énonciation en linguistique à base de corpus s'explique par le passage d'une linguistique de la langue à une linguistique du « discours » donc, un changement de l'objet même de la linguistique à partir des années 1960. En effet, la prise en compte de tous les phénomènes liés aux conditions de production du discours apparaît comme pertinente pour la compréhension du fonctionnement de la langue.

Lorsqu'on aborde le sens des unités linguistiques dans un corpus, on est inévitablement amené à les relier à des facteurs extralinguistiques, c'est-à- dire à leur référence comme à leur prise en charge par un énonciateur.

L'approche énonciative trouve son fondement dans les travaux d'Émile Benveniste (1966, 1970), qui définit l'énonciation comme la **mise en fonctionnement de la langue par un acte individuel d'utilisation**. Autrement dit, la langue n'existe réellement que lorsqu'un locuteur s'en empare pour produire un discours dans une situation donnée.

Benveniste insiste sur le fait que chaque énonciation est un acte unique, situé, et dépendant d'un sujet parlant. Il s'agit donc d'un processus d'appropriation par lequel le locuteur mobilise les ressources de la langue pour exprimer sa position et entrer en relation avec un interlocuteur.

Ainsi, avant l'énonciation, la langue est une potentialité ; après l'énonciation, elle devient une réalité discursive inscrite dans un contexte précis. L'énonciation introduit donc nécessairement :

- un locuteur,
- un allocutaire,
- un rapport au monde (référence),
- une situation spatio-temporelle.

2. L'appareil formel de l'énonciation

Benveniste met en évidence l'existence d'un appareil formel de l'énonciation, constitué d'indices linguistiques permettant d'inscrire le locuteur dans son discours.

Ainsi, dans l'idée de l'auteur, il suffirait à un analyste de porter le regard sur l'acte par lequel le discours est produit pour se rendre compte que le locuteur est le paramètre essentiel dans la mise en fonctionnement de la langue.

Voici le texte intégral d'Émile Benveniste, « L'appareil formel de l'énonciation », *Langages*, 17/1970, Paris, Didier-Larousse

L'acte individuel par lequel on utilise la langue introduit d'abord le locuteur comme paramètre dans les conditions nécessaires à l'énonciation. Avant l'énonciation, la langue n'est que possibilité de la langue. Après l'énonciation, la langue est effectuée en une instance de discours, qui émane d'un locuteur, forme sonore qui atteint un auditeur et qui suscite une autre énonciation en retour.
En tant que réalisation individuelle, l'énonciation peut se définir, par rapport à la langue, comme un processus d'appropriation. Le locuteur s'approprie l'appareil formel de la langue et il énonce sa position de locuteur par des indices spécifiques (...)
Mais immédiatement, dès qu'il se déclare locuteur et assume la langue, il implante l'autre en face de lui, quel que soit le degré de présence qu'il attribue à cet autre. Toute énonciation est, explicitement ou implicitement, une allocution, elle postule un allocutaire.
Enfin, dans l'énonciation, la langue se trouve employée à l'expression d'un certain rapport au monde (...) La référence est la partie intégrante de l'énonciation.
L'acte individuel d'appropriation de la langue introduit celui qui parle dans sa parole. C'est là une des données constitutives de l'énonciation. La présence du locuteur à son énonciation fait que chaque instance de discours constitue un centre de référence interne. Cette situation va se manifester par un jeu de formes spécifiques dont la fonction est de mettre le locuteur en relation constante et nécessaire avec son énonciation.
Cette description un peu abstraite s'applique un phénomène linguistique familier dans l'usage, mais dont l'analyse théorique commence seulement. C'est d'abord l'émergence des indices de personne (le rapport je-tu), qui ne se produit que dans et par l'énonciation : le terme je dénotant l'individu qui profère l'énonciation, le terme tu l'individu qui y est présent comme allocutaire.
De même nature et se rapportant à la même structure d'énonciation sont les indices nombreux d'ostension (type ce, ici, etc.), termes qui impliquent un geste désignant l'objet en même temps qu'est prononcée l'instance du terme.
Les formes appelées traditionnellement « pronoms personnels », « démonstratifs » nous apparaissent maintenant comme une classe d' « individus linguistiques », qu'il s'agisse des personnes, des moments, des lieux, par opposition aux termes nominaux qui renvoient toujours seulement à des concepts. Or le statut de ces « individus linguistiques » tient au fait qu'ils naissent d'une énonciation, qu'ils sont produits par cet événement individuel (...). Ils sont engendrés à nouveau chaque fois qu'une énonciation est proférée, et chaque fois ils désignent à neuf.
Une troisième série de termes afférents à l'énonciation est constituée par le paradigme entier - souvent vaste et complexe - des formes temporelles, qui se déterminent par rapport à l'EGO, centre de l'énonciation. Les « temps » verbaux dont la forme axiale, le présent, coïncide avec le moment de l'énonciation, font partie de cet appareil nécessaire.
La temporalité (...) est produite dans et par l'énonciation. De l'énonciation procède l'instauration de la catégorie du présent, et de la catégorie du présent naît la catégorie du temps. Le présent est proprement la source du temps. Il est cette présence au monde que l'acte de l'énonciation rend seul possible, car, qu'on veuille bien y réfléchir, l'homme ne dispose d'aucun autre moyen de vivre le « maintenant » et de le faire actuel que de le réaliser par l'insertion du discours dans le monde. On pourrait montrer par des analyses des systèmes temporels en diverses langues la position centrale du présent. Le présent formel ne fait qu'expliciter le présent inhérent à l'énonciation, qui se renouvelle avec chaque production de discours, et à partir de ce présent continu, coextensif à notre présence propre, s'imprime dans la conscience le sentiment d'une continuité que nous appelons « temps » : continuité et temporalité s'engendrant dans le présent incessant de l'énonciation qui est le présent de l'être même, et se délimitant, par référence interne, entre ce qui va devenir présent et ce qui vient de n'être plus (...)

Il faut donc distinguer les entités qui ont donc dans la langue leur statut plein et permanent et celles qui, émanant de l'énonciation, n'existent que dans le réseau d' « individus » que l'énonciation crée et par rapport l' « ici-maintenant » du locuteur (...). L'énonciation donne les conditions nécessaires aux grandes fonctions syntaxiques.
Dès lors que l'énonciation se sert de la langue pour influencer en quelque manière le comportement de l'allocutaire, il dispose à cette fin d'un appareil de fonctions. C'est, d'abord, l'interrogation, qui est une énonciation construite pour susciter une « réponse » (...)
On y attribuera pareillement les formes que nous appelons d'intimation : ordres, appels conçues dans des catégories comme l'impératif, le vocatif impliquant un rapport vivant et immédiat de l'énonciateur à l'autre dans une référence immédiate au temps de l'énonciation.
Dans son tour syntaxique comme dans son intonation, l'assertion vise à communiquer une certitude, elle est la manifestation la plus commune de la présence du locuteur dans l'énonciation, elle a même des instruments spécifiques qui l'expriment ou l'impliquent (...).
Se rangent ici toutes sortes de modalités formelles, les unes appartenant aux verbes comme les « modes » (optatif, subjonctif) énonçant les attitudes de l'énonciateur à l'égard de ce qu'il énonce (attente, souhait, appréhension), les autres à la phraséologie (« peut-être », « sans doute », « probablement ») et indiquant incertitude, possibilité, indécision, etc.
Ce qui en général caractérise l'énonciation est l'accentuation de la relation discursive au partenaire, que celui-ci soit réel ou imaginé, individuel ou collectif (...) Comme forme de discours, l'énonciation pose deux « figures », également nécessaires, l'une source, l'autre but de l'énonciation. C'est la structure du dialogue. Deux figures en position de partenaires sont alternativement protagonistes de l'énonciation. (...) Le monologue procède bien de l'énonciation. Le « monologue » est un dialogue intériorisé, formulé en « langage intérieur », entre un moi locuteur et un moi écouteur. Parfois le moi locuteur est le seul à parler : le moi écouteur reste néanmoins présent ; sa présence est nécessaire et suffisante pour rendre signifiante l'énonciation du moi locuteur (...)

2.1. Les pronoms personnels (embrayeurs)

Les pronoms personnels (je, tu, nous, vous) sont des marqueurs fondamentaux de l'énonciation. Ils désignent les participants à l'acte de parole :

- je : le locuteur,
- tu : l'allocutaire,
- il : la non-personne (objet du discours).

Ces éléments sont appelés *embrayeurs*, car ils relient l'énoncé à la situation d'énonciation.

2.2. Les EMBRAYEURS et les déictiques

Les déterminants (mon, ton, ce…) et les indicateurs spatiaux et temporels (ici, maintenant, aujourd'hui) organisent le discours autour du locuteur et de sa situation.

2.3. Les temps verbaux

Benveniste distingue deux systèmes temporels :

- Le discours : centré sur le présent d'énonciation (présent, passé composé, futur).
- Le récit : détaché du moment d'énonciation (passé simple, imparfait).

Cette distinction montre que la temporalité est construite dans et par l'énonciation.

Déictiques	
Emploi de déictiques › Mots ne pouvant être compris que si l'on connaît la situation d'énonciation. • Marques de 1re et 2e pers. → *je, tu, vous, mon, votre, le mien, le nôtre, toi, etc.* • Indicateurs spatiaux et temporels en référence à la situation d'énonciation → *ici, là-bas, à droite, aujourd'hui, hier, etc.*	Absence de déictiques • Marques de 3e pers. → *il, elles, son, ses, les leurs, etc.* • Indicateurs spatiaux et temporels sans référence à la situation d'énonciation → *à Paris, au 108 rue Balzac, ce jour-là, etc.*
Modalisateurs	
Emploi de modalisateurs › Éléments qui expriment une subjectivité du locuteur → *à mon avis, sans doute, magnifique ≠ horrible, etc.* → *Oui, tu as raison, c'est une riche idée.*	Absence de modalisateurs → *Le tramway de Neuilly venait de passer la porte Maillot et il filait […] tout le long de la grande avenue qui aboutit à la Seine.*

3. Énoncé, énonciation et situation d'énonciation

Il convient de distinguer trois notions fondamentales :

- L'énoncé : le message produit (oral ou écrit).
- L'énonciation : l'acte de production de ce message.
- La situation d'énonciation : l'ensemble des conditions dans lesquelles le message est produit.

Analyser une situation d'énonciation revient à répondre aux questions :

- qui parle ?
- à qui ?
- où ?
- quand ?
- pourquoi ?

Exemple

« Il est 14 heures, nous sommes en salle 102 pour le cours de sociolinguistique. »

Cet énoncé contient des indices clairs de la situation d'énonciation (temps, lieu, participants).

4. Les modes d'énonciation : énoncé ancré et énoncé coupé

4.1. L'énoncé ancré dans la situation d'énonciation

Dans ce cas, le locuteur est présent dans son discours, et celui-ci contient des marques de subjectivité :

- pronoms de première et deuxième personne,
- déictiques,
- modalisateurs.

Exemple :

« Oui, tu as raison, c'est une excellente idée. » La compréhension dépend du contexte.

4.2. L'énoncé coupé de la situation d'énonciation

Le locuteur s'efface et le discours devient autonome :

- absence de déictiques,
- prédominance de la troisième personne,
- objectivité apparente.

Exemple :« Le tramway venait de passer la porte Maillot… »

L'énoncé est compréhensible sans référence à la situation d'énonciation.

Énoncé ancré Avec marques de la situation d'énonciation	Énoncé coupé Sans marques de la situation d'énonciation
Genres principaux	
Dialogues, lettres, articles de presse, journaux intimes	Romans, contes, récits historiques, textes scientifiques
Temps principaux	
Présent de l'indicatif, passé composé, futur, présent de l'impératif	Passé simple, imparfait, plus-que-parfait, présent de narration ou de vérité générale

Le système énonciatif distingue entre discours et récit.

Choisir de se manifester et d'inscrire sa présence dans l'énoncé : c'est le mode du discours. Ne pas confondre ce sens du mot discours avec son sens oratoire de texte prononcé devant un public.

Il est ancré dans la situation d'énonciation, il s'inscrit dans le contexte présent du texte ou de la parole : l'énonciateur commente alors un récit, exprime des idées ou des sentiments ; il peut aussi solliciter le lecteur, l'interpeller (*injonctions à l'impératif, interrogations oratoires*…).

L'énonciateur et le destinataire sont donc présents dans le texte sous forme de pronoms et déterminants de *première et deuxième personne.*

Le discours peut se rencontrer :

- à l'**écrit** (dans la morale des fables, dans des poèmes, des autobiographies, des articles de presse…)

Ex : « Je sens en écrivant ceci que mon pouls s'élève encore ; ces moments me seront toujours présents quand je vivrais cent mille ans.» Jean-Jacques Rousseau, *Confessions* (1778).

* Rousseau inscrit son texte dans le présent de l'énonciation : *« je sens »* et se projette dans l'avenir : *« ces moments me seront toujours présents…» ; le* pronom démonstratif *« ceci »* désigne ce qu'il est en train d'écrire.

- à l'**oral** (dans les genres oratoires).

Ex : « La misère, messieurs, j'aborde ici le vif de la question, voulez-vous savoir jusqu'où elle est, la misère ? Voulez-vous savoir jusqu'où elle peut aller, jusqu'où elle va, je ne dis pas en Irlande, je ne dis pas au Moyen Âge, je dis en France, je dis à Paris, et au temps où nous vivons ?» Victor Hugo : « Détruire la misère » (Discours à l'Assemblée Nationale du 9 juillet 1849).

* Les indications de temps et de lieu se réfèrent au moment et au lieu de l'écriture (ou de la parole) : *« ici » ; « en France, à Paris, au temps où nous vivons »*.

* Les temps des verbes s'organisent par rapport au présent de l'énonciation : *« j'aborde ici »*

* L'orateur s'adresse directement à l'auditoire : *« Voulez-vous savoir… ? »*

Pour le récit

Le locuteur peut choisir de rester absent, de n'établir aucune relation directe avec le destinataire : c'est le mode du récit. Ne pas confondre récit et narration. Une narration peut être un récit si le narrateur choisit de s'effacer, mais elle devient un discours si le narrateur choisit de donner dans l'énoncé des indices de sa présence.

Il est **coupé de la situation d'énonciation** : des faits, des actions sont rapportés, mais l'énonciateur (le narrateur) s'efface dans l'énoncé. Le destinataire (le lecteur), également, n'est pas explicitement sollicité dans le texte.

On rencontre des récits dans les romans, les contes, les nouvelles, les biographies, les autobiographies, les fables…

Ex : « Bertin, après son retour de Rome, en 1864, était demeuré quelques années sans succès et sans renom ; puis soudain, en 1868, il exposa sa *Cléopâtre* et fut en quelques jours porté aux nues par la critique et le public. » Guy de Maupassant, *Fort comme la mort* (1889).

* Les pronoms et déterminants sont ceux de la troisième personne : *il, sa, son.*

* Les indications de temps et de lieu, et les temps verbaux, se réfèrent au contexte des événements racontés, qui appartiennent au passé : *en 1864, en 1868, en quelques jours… ; plus-que-parfait, passé simple.*

	Le récit	**Le discours**
Les temps dominants	Le système des temps verbaux renvoie à un passé qui n'est pas situé par rapport au moment de l'énonciation, d'où l'emploi du passé simple ou du présent de narration ou de l'imparfait, plus-que-parfait.	Le système des temps situe l'énoncé par rapport au moment de l'énonciation, d'où l'emploi fréquent du passé composé quand le locuteur renvoie au passé, et la prédominance du présent et

		parfois du futur.
Les pronoms personnels	Le locuteur et l'énonciateur s'effacent au profit de la troisième personne mais si l'on rencontre le – je -, il s'agit du locuteur se prenant lui-même comme objet (autobiographie), ou d'un narrateur fictif distinct de l'auteur.	La première et la deuxième personnes sont utilisées dès que s'affirme la relation entre le locuteur et le destinataire.
Les indicateurs de temps et de lieu	Le repérage se fait par rapport aux événements entre eux; ce sont des lieux et des moments internes à l'histoire. Par exemple : *La veille, trois mètres plus loin.*	Le repérage se fait par rapport à la situation d'énonciation, c'est- à-dire par rapport au présent du locuteur et au lieu qu'il occupe. Par exemple : *Hier, à trois mètres d'ici*.
Les indicateurs du degré de conviction et de l'opinion du locuteur	Ces révélateurs sont absents, dans la mesure où le locuteur s'efface et ne fait pas intervenir sa façon de voir.	Le locuteur est présent ; il prend position quant à la vérité ou la fausseté et la certitude ou l'incertitude de son propre énoncé.

Conclusion

L'approche énonciative constitue une étape essentielle dans l'évolution de l'analyse linguistique, en mettant en avant le sujet parlant et les conditions de production du discours. Elle permet de comprendre que le langage n'est pas seulement un système de signes, mais un acte social et interactionnel, inscrit dans une situation précise.

En articulant les dimensions linguistiques et extralinguistiques, cette approche offre des outils puissants pour analyser les discours dans leur complexité.

L'énonciation permet également d'exprimer l'attitude du locuteur à travers les modalités (certitude, doute), les expressions modalisatrices (peut-être, sans doute) et la subjectivité qui traduisent l'engagement du locuteur dans son discours. Elle s'avère particulièrement pertinente dans l'étude des corpus contemporains, notamment en contexte pédagogique, médiatique ou numérique, où la subjectivité, l'interaction et le contexte jouent un rôle déterminant dans la construction du sens.

COURS 18: L'APPROCHE LEXICOMETRIQUE

INTRODUCTION

Avec le développement des technologies numériques, l'analyse du discours a connu un tournant méthodologique majeur grâce à l'émergence de la lexicométrie. Cette approche, à l'intersection de la linguistique, de la statistique et de l'informatique, permet d'explorer de vastes corpus textuels à partir de traitements automatisés.

Elle répond à un besoin croissant d'objectivation des données linguistiques, en fournissant des outils pour mesurer, comparer et modéliser les usages du langage.

1. Définition et évolution de la lexicométrie

La lexicométrie est une méthode d'analyse quantitative des textes reposant sur des outils informatiques. Elle est également appelée :

- analyse du discours assistée par ordinateur,
- traitement automatisé du discours.

La lexicométrie est une approche fusionnant linguistique, statistique et informatique. Cette jeune discipline est appelée également « analyse du discours assistée par ordinateur », ou encore « traitement automatisé du discours », « l'étude scientifique du discours faite avec l'outil informatique ». Elle se consacre à l'analyse quantitative des textes, explorant les données linguistiques, sémantiques et stylistiques au sein de vastes corpus.

Dans les années 1970/80, cette approche a prospéré avec l'avènement de la micro-informatique. Le Groupe de Saint-Cloud en France a joué un rôle clé, facilitant le traitement et l'analyse de textes volumineux. La revue "MOTS – Mots, Ordinateurs, Textes, Sociétés," fondée en 1980, a été un vecteur essentiel pour ses applications. Parmi ses principaux fondateurs :

- Charles Muller (lexicométrie littéraire, a exploré le style et l'évolution du vocabulaire des écrivains.),
- Maurice Tournier (lexicologie politique, a examiné les liens entre langage politique et pouvoir.).

Elle permet de répondre à plusieurs questions relatives à :

-L'usage des termes dans le corpus

Comment les mots sont-ils employés et dans quels contextes ?

-Évolution du vocabulaire

Comment le lexique évolue-t-il dans le temps ou selon les auteurs ?

-Fréquences , distribution du lexique du corpus

Quels sont les mots les plus fréquents ?

-Thèmes du corpus, Quels sujets dominent ?

- Relations entre mots et réseaux sémantiques

Quelles associations lexicales sont récurrentes ?

2. Outils et méthodes d'analyse

Les logiciels utilisés incluent :

- Tropes,
- Hyperbase,
- Alceste,
- Sphinx,
- Prospero.

La méthode repose sur :

- le repérage des mots (items graphiques),
- l'analyse statistique,
- l'étude des relations lexicales (cooccurrences).

Conclusion

La lexicométrie constitue un outil puissant pour l'analyse des grands corpus, en apportant une dimension quantitative et objectivable à l'étude du discours. Toutefois, ses résultats doivent être interprétés qualitativement pour en saisir toute la portée. Elle s'inscrit ainsi dans une approche mixte, combinant rigueur statistique et analyse contextualisée.

La lexicométrie s'appuie sur des traitements statistiques et informatiques. Elle choisit des items formels (mots graphiques), crée des relations statistiques sur des axes paradigmatique, syntagmatique et situationnel, avec une norme intrinsèque définie à l'intérieur des corpus.

COURS 19 : L'APPROCHE DIALOGIQUE ET SEMIOLOGIQUE

Introduction

Au-delà des approches structurales et quantitatives, qui privilégient l'étude des formes linguistiques et des régularités statistiques, certaines perspectives contemporaines mettent l'accent sur la dimension interprétative, interactionnelle et symbolique du langage. Dans cette optique, le discours n'est plus considéré comme un simple assemblage d'unités linguistiques, mais comme un espace dynamique où circulent des voix, des significations et des représentations.

L'approche dialogique et l'approche sémiologique s'inscrivent pleinement dans cette orientation. Elles permettent d'explorer le discours en tant que phénomène social et culturel, en mettant en lumière, d'une part, la pluralité des voix qui s'y manifestent, et d'autre part, les systèmes de signes à travers lesquels le sens se construit, notamment dans les images et les supports médiatiques.

1. L'approche dialogique : le langage comme interaction

1.1. Fondements du dialogisme

Développé par le penseur russe Mikhaïl Bakhtine, le dialogisme repose sur l'idée fondamentale que tout discours est intrinsèquement pluriel. Aucun énoncé n'est isolé : il est toujours en relation avec d'autres discours, antérieurs ou contemporains. Cette conception introduit la notion de **polyphonie**, selon laquelle plusieurs voix, points de vue ou consciences coexistent au sein d'un même énoncé.

Dans cette perspective, le langage est fondamentalement interactionnel : parler, c'est toujours répondre à d'autres discours et anticiper des réactions. Le sens émerge ainsi de cette interaction permanente entre les voix.

1.2. Objectifs de l'approche dialogique

L'analyse dialogique vise plusieurs objectifs essentiels :

- **Révéler la diversité des voix** : identifier les différentes instances discursives présentes dans un énoncé ;
- **Analyser les interactions** : comprendre comment ces voix dialoguent, se complètent ou s'opposent ;

- **Comprendre les dynamiques discursives** : mettre en évidence les phénomènes d'accord, de tension ou de conflit entre les points de vue.

Cette approche permet ainsi de saisir la complexité du discours en tant que lieu de confrontation et de négociation du sens.

1.3. Méthode d'analyse dialogique

L'analyse dialogique repose sur plusieurs étapes complémentaires :

- **Identification des voix** : repérer les différentes sources discursives (discours rapporté, implicite, intertextuel) ;
- **Analyse des dialogues internes** : examiner les interactions entre ces voix au sein du discours ;
- **Étude de l'intertextualité** : analyser les liens entre le discours étudié et d'autres discours (références culturelles, idéologiques, médiatiques).

Exemple **:** Dans un débat en ligne (par exemple sur les réseaux sociaux), on observe :

- la coexistence de discours politiques, médiatiques la confrontation d'opinions divergentes ;
- des références implicites à d'autres discours (articles, vidéos, slogans). L'approche dialogique permet ici de comprendre comment ces différentes voix interagissent et construisent le sens collectif du débat.

2. L'approche sémiologique : analyse des signes et des images

2.1. Fondements de la sémiologie

L'approche sémiologique, issue des travaux de Ferdinand de Saussure et développée par Roland Barthes, se consacre à l'étude des systèmes de signes dans la communication humaine.

Elle ne se limite pas au langage verbal, mais s'étend à l'ensemble des formes de signification, notamment les images, les symboles et les objets culturels. Dans cette perspective, toute production (texte, image, publicité) est envisagée comme un ensemble de signes porteurs de sens.

2.2. Les niveaux de signification selon Barthes

Roland Barthes distingue trois niveaux de signification dans toute image :

Le message linguistique : les éléments textuels (slogan, titre, légende) qui orientent l'interprétation ;

Le message iconique dénoté : la signification littérale, ce que l'on voit objectivement ;

Le message iconique connoté : les significations implicites, symboliques ou culturelles associées à l'image.

2.3. Concepts clés : dénotation et connotation

- **Dénotation** : correspond au sens objectif, immédiat et descriptif d'un élément visuel.
- **Connotation** : elle renvoie aux significations implicites, culturelles et subjectives, qui varient selon les contextes et les interprétations.

Exemple : Prenons une publicité pour une voiture :

- **Dénotation** : une voiture sur une route.
- **Connotation** : liberté, réussite sociale, modernité, puissance. L'image ne se contente pas de montrer un objet, elle véhicule un ensemble de valeurs symboliques.

3. Analyse sémiologique de l'image

L'analyse sémiologique vise à comprendre comment les images produisent du sens à travers différents paramètres :

a) **les couleurs** : associées à des valeurs symboliques (rouge = passion, danger, énergie) ;
b) **la composition** : organisation des éléments visuels dans l'espace ;
c) **les symboles** : signes culturels porteurs de significations ;
d) **le contexte culturel** : cadre social et historique de production et de réception ;
e) **l'intention de l'auteur** : objectif communicatif ou stratégique.

Cette analyse permet de dépasser l'apparence immédiate de l'image pour en révéler les significations profondes.

Conclusion

Les approches dialogique et sémiologique apportent un éclairage complémentaire et indispensable à l'analyse du discours. Tandis que le dialogisme met en évidence la dimension interactive et polyphonique du langage, la sémiologie permet d'explorer les mécanismes de production du sens à travers les signes, notamment visuels.

Ensemble, ces approches permettent de dépasser une vision strictement linguistique du discours pour intégrer les dimensions sociales, culturelles et idéologiques qui le structurent. Elles s'avèrent particulièrement pertinentes pour l'étude des discours contemporains, notamment dans les environnements numériques et médiatiques, où se multiplient les interactions et les formes hybrides de communication.

PARTIE II

TRAVAUX PRATIQUES.

« Plongez au cœur des langues à travers les données »



OBJECTIFS SPÉCIFIQUES DES TRAVAUX DIRIGÉS (TD)

Les travaux pratiques de ce module visent à développer chez l'étudiant une véritable compétence de terrain en linguistique de corpus, en articulant rigueur méthodologique, autonomie de recherche et capacité d'analyse empirique des données langagières. Ils permettent d'opérationnaliser les savoirs théoriques acquis en cours et de former progressivement l'étudiant à la démarche du chercheur.

Dans cette perspective, les objectifs spécifiques sont les suivants :

1. **Savoir concevoir un corpus en lien direct avec une problématique de recherche précise**

   Il s'agit d'apprendre à délimiter un objet d'étude, à définir des critères de sélection pertinents et à construire un corpus cohérent, représentatif et scientifiquement exploitable.

2. **Maîtriser les techniques de collecte des données linguistiques (orales et écrites) ainsi que des métadonnées associées**

L'étudiant sera amené à identifier, sélectionner et recueillir des données authentiques issues de différents contextes de communication, tout en intégrant les informations contextuelles nécessaires (auteur, date, genre, support, situation de communication, etc.).

3. **Développer des compétences en transcription, organisation et structuration des corpus**

   Cette étape consiste à apprendre à transformer des données brutes en un corpus exploitable, à travers des opérations de transcription, de classification, de nettoyage et de tri méthodologique des données.

4. **S'initier à la formulation des problématiques et à la construction d'hypothèses de recherche**

Les étudiants seront formés à transformer des observations empiriques en questions scientifiques pertinentes, puis à formuler des hypothèses claires, vérifiables et adaptées au corpus constitué.

5. **S'exercer au traitement et à l'analyse de corpus authentiques issus de l'ASJP (Algerian Scientific Journal Platform)**

   Les étudiants seront initiés à l'exploitation de données scientifiques réelles, extraites de revues académiques, afin d'analyser des phénomènes linguistiques dans des contextes de production authentiques et variés.

Progression méthodologique des travaux dirigés en linguistique de corpus

Les travaux dirigés du cours de linguistique de corpus sont organisés de manière progressive afin d'accompagner l'étudiant dans toutes les étapes d'une recherche scientifique empirique. Ils débutent par le **TD01**, consacré à la formulation de la problématique, étape fondamentale qui oriente l'ensemble du travail de recherche, puis se poursuivent avec le **TD02**, qui présente des exemples de recherches empiriques à base de corpus afin d'illustrer concrètement les démarches méthodologiques. Le **TD03** aborde la construction des hypothèses de recherche, tandis que le **TD04** propose une réflexion sur la définition du corpus en analyse conversationnelle. Les **TD05 à TD07** sont centrés sur les modalités de constitution des données, en mettant l'accent sur les critères de recueil, l'observation en général et l'observation en linguistique de terrain.

Dans une perspective comparative, le **TD09** traite des méthodes d'entretien et de questionnaire, complété par le **TD10** qui présente les outils et techniques de collecte des données en linguistique. L'analyse et l'interprétation des données sont ensuite développées dans le **TD11**, suivies du **TD12**, qui propose l'étude de corpus variés à partir d'extraits de l'ASJP, et du **TD13**, consacré aux outils d'analyse tels que l'analyse de contenu. Les **TD14 et TD15** mettent en relation thématique et choix du corpus à travers des études de cas, permettant de comprendre l'adéquation entre objet de recherche et données. Cette réflexion est prolongée dans le **TD16**, qui approfondit l'étude de corpus diversifiés. Enfin, le **TD17** ouvre sur une perspective appliquée en montrant le passage du mémoire académique au mémoire de type start-up, en insistant sur l'adaptation des corpus aux objectifs professionnels.

Ainsi, l'ensemble de ces travaux dirigés vise à doter l'étudiant d'une maîtrise méthodologique complète, allant de la conception du projet de recherche à l'exploitation et à la valorisation des données de corpus.

TD 01 : ETAPES D'UNE RECHERCHE EMPIRIQUE, A BASE DE CORPUS : La problématique

ACTIVITE 01 : Pour chaque thématique proposée, ici, dans le domaine de la linguistique, choisissez la problématique la plus pertinente **:**

La prononciation du français en Algérie

- Quelles sont les caractéristiques phonologiques des variétés dialectales de l'Est algérien ?
- Quelles sont les causes de la prononciation défaillante des voyelles oralo-nasales du français chez les algériens ?
- Quelles sont les étapes de l'acquisition du langage parlé chez les enfants ?

L'étude sociolinguistique des langues en Algérie.

- Quelles sont les formes d'alternances codiques (arabe/français) les plus répandues et leurs fonctions dans les pratiques langagières des algériens ?
- Quel est le statut de la langue française, face à l'Arabe, et à l'anglais à l'Université Algérienne ?
- Quelles sont les marques de subjectivité et les stratégies discursives utilisées par les journalistes de « Liberté » ?

ACTIVITE 02 : Voici une liste de thématiques, en domaine de la linguistique. Quelle problématique proposez-vous pour ces sujets de recherche?

a) Etude comparative des expressions figées de la langue française dans la presse francophone et algérienne
   Problématique : ……
b) Etude sociolinguistique des affiches publicitaires dans les rues de la ville d'ORAN. (ou autre).
c) Analyse discursive des titres journalistiques dans la presse francophone écrite pendant le Hirak :Cas du journal…. (Au choix)
d) Etude onomastique des toponymes de la région de Zemmoura. (ou autre)
e) L'emprunt lexical dans le domaine commercial
f) Les néologismes et la création lexicale pendant la période de Pandémie du COVID 19.
g) Les phénomènes de création lexicale chez les Youtubeurs algériens.
h) L'impact des représentations de la langue française sur l'enseignement/apprentissage de cette langue. Problématique : ……

ACTIVITE 03 : Voici une liste de thématiques, en domaine de la linguistique. Il faut les personnaliser, c'est-à-dire les transformer en une problématique pertinente pour qu'il soit adapté à une recherche empirique que vous souhaiteriez faire.

a) Etude morphosyntaxique des écrits des panneaux publicitaires de la ville .
b) Etude de la structure narrative de quelques contes populaires de la région de l'Est algérien.
c) Le statut de la langue française, face à l'Anglais, dans les filières scientifiques à l'Université algérienne.
d) Lecture sociolinguistique de la trame romanesque de Malika MOKEDDEM.
e) Etude comparative des proverbes et dictons français et algérien sur le thème de « la femme »
f) Étude diachronique de l'anthroponymie à Bordj Bou Arréridj (1980–2020)
g) Étude linguistico-pragmatique du discours politique de Macron et de D. Trump autour de la reconnaissance de l'État de Palestine.

COMPTE RENDU

ACTIVITÉ 01 : Choix de la problématique pertinente

1. La prononciation du français en Algérie :

→ Problématique retenue :

Quelles sont les causes de la prononciation défaillante des voyelles oralo-nasales du français chez les Algériens ?

✓ Plus directement liée à la thématique (prononciation du français).

2. L'étude sociolinguistique des langues en Algérie :

→ Problématique retenue :

*Quelles sont les formes d'alternances codiques (arabe/français) les plus répandues et leurs fonctions dans les pratiques langagières des Algériens ?*Pertinente pour une approche sociolinguistique des usages.

ACTIVITÉ 02 : Propositions de problématiques

a) Expressions figées dans la presse→ *Quelles différences d'usage et de fonction des expressions figées observe-t-on entre la presse francophone française et algérienne ?*

b) Affiches publicitaires à Oran→ *Comment les choix linguistiques (langues, alternance codique, registres) dans les affiches publicitaires à Oran reflètent-ils les pratiques sociolinguistiques locales ?*

c) Titres journalistiques pendant le Hirak→ *Quelles stratégies discursives et idéologiques se manifestent dans les titres de presse francophone durant le Hirak ?*

d) Toponymes de Zemmoura→ *Quels sont les processus de formation et les significations socioculturelles des toponymes de la région de Zemmoura ?*

e) Emprunt lexical commercial→ *Quelles sont les fonctions et les motivations des emprunts lexicaux dans le discours commercial en Algérie ?*

f) Néologismes (COVID-19)→ *Quels types de néologismes ont émergé durant la pandémie de COVID-19 et quelles fonctions remplissent-ils dans le discours social ?*

g) Création lexicale chez les Youtubeurs→ *Quels procédés de création lexicale utilisent les Youtubeurs algériens et quelles sont leurs fonctions identitaires et expressives ?*

h) Représentations du français et apprentissage→ *Comment les représentations sociales de la langue française influencent-elles son enseignement/apprentissage en Algérie ?*

ACTIVITÉ 03 : Transformation des thématiques proposées en problématiques empiriques

a) Panneaux publicitaires→ *Quelles structures morphosyntaxiques caractérisent les écrits des panneaux publicitaires dans la ville de X, et quelles fonctions communicatives remplissent-elles ?*

b) Contes populaires (Est algérien)→ *Quelles structures narratives dominent dans les contes populaires de l'Est algérien et quelles valeurs culturelles véhiculent-elles ?*

c) Français vs anglais (filières scientifiques)→ *Quel est le statut et l'usage effectif du français face à l'anglais dans les filières scientifiques à l'université X ?*

d) Œuvre de Malika Mokeddem→ *Comment les dynamiques sociolinguistiques se manifestent-elles dans la trame romanesque des œuvres de Malika Mokeddem ?*

e) Proverbes sur la femme→ *Quelles représentations de la femme émergent dans les proverbes français et algériens et en quoi diffèrent-elles culturellement ?*

f) Étude diachronique de l'anthroponymie à Bordj Bou Arréridj (1980–2020)
Problématique : Comment les choix anthroponymiques (prénoms, patronymes, formes

d'adresse) ont-ils évolué à B. B. Arréridj entre 1980 et 2020 et quelles relations ces changements entretiennent-ils avec urbanisation, mobilité et politiques identitaires ?

g) Étude linguistico-pragmatique du discours politique de Macron et de D. Trump autour de la reconnaissance de l'État de Palestine.

. Problématique : Quelles stratégies argumentatives sont déployées par Macron et Trump dans leur allocutions officielles autour de la Palestine ?

TD 02 : QUELQUES RECHERCHES EMPIRIQUES A BASE DE CORPUS

ACTIVITE :

a) -Identifiez l'objectif de chaque « recherche » proposée dans chaque extrait ci-dessous.

b) Identifiez à partir des résumés proposés, les « niveaux de recherche » de chaque thématique.

c) Quelle recherche vous semble, la plus intéressante et pourquoi ?

1/Emprunt et alternance codique dans la chanson raï et rap en Algérie

Résumé : Cette recherche se donne comme principale tâche de décrire et d'expliquer la présence du français dans la production artistique des jeunes chanteurs de raï et de rap, pour savoir si ce mélange linguistique – qui se présente parfois sous forme d'emprunt et parfois sous forme d'alternance codique – est la conséquence du bilinguisme ou multilinguisme qui caractérise la société algérienne ou si c'est une nouvelle variation liée à d'autres facteurs sociaux. A partir d'un corpus choisi, nous tenterons de répondre à quelques questions : Quelle est la raison qui pousse les jeunes chanteurs à produire des textes bilingues (arabe, français) ? Pourquoi les paroliers préfèrent des mots français comme (l'Amour, l' visa, l'aéroport) au vocabulaire relatif aux mêmes thèmes dans l'arabe classique ?Le recours à l'emprunt au français répond-il à une réelle insuffisance de la langue arabe, ou plutôt à la demande d'un public qui semble être « francophone » ? Nous avons choisi ces deux styles musicaux (raï et rap) comme notre champ d'investigation car ils paraissent les plus représentatifs dans une Algérie multiple où chaque région possède son genre musical, mais où le raï et le rap sont écoutés à travers le pays et par toutes les catégories d'âge et de sexe.

2/ *Ironie Et Argumentation dans Le Discours Journalistique Algérien. Cas Des Chroniques Points Zéro Et Pousse Avec Eux*

Résumé :De par sa nature multidimensionnelle, l'ironie met à la disposition du journaliste toute une stratégie discursive dont le fonctionnement est assez complexe. À cet égard, des recherches récentes (Berrendonner, 1981, 2002 ; Eggs, 2009 ; Perrin, 1996) montrent que l'aspect argumentatif de l'ironie dans la construction du discours est inébranlable. S'inscrivent dans cette perspective, tout en s'appuyant sur la théorie récente de l'ironie proposée par Eggs (2009), cette recherche se propose de mettre en évidence le rôle argumentatif de l'ironie dans un genre journalistique particulier à savoir la chronique. Notre objectif dans cette étude est double. D'une part, nous allons voir quels sont les mécanismes discursifs de l'ironie auxquels fait appel le journaliste pour construire son discours argumentatif. D'autre part, il s'agit de tenter de faire comprendre la place accordée à l'usage argumentatif de l'ironie dans ce genre de discours pour dire l'événement.

3/Le pseudonyme chez la communauté d'internautes algériens sur le web social : quels choix et quels enjeux sociolinguistiques ?

Le pseudonyme comme catégorie anthroponymique est une pratique sociale dont l'intérêt que lui portent les chercheurs dans le champ de l'onomastique populaire et de la sociolinguistique est quasi minime d'où, notamment, la rareté des recherches qui lui ont été consacrées. Cette contribution vient justement, à contrecourant de ce regain d'intérêt manifeste pour cette catégorie anthroponomastique. Elle porte sur les choix et enjeux sociolinguistiques du pseudonyme d'une communauté d'internaute. La problématique principale est de déterminer quels motivations et critères sociolinguistiques, identitaires…présidant aux choix des pseudonymes chez la communauté algérienne et quels usages en fait cette communauté du pseudonyme et dans quels contextes. A travers l'analyse socio-onomastique et l'exploitation des discours dans une perspective sociolinguistique, que le choix du pseudonyme se fait dans des registres onomastiques et linguistiques divers. Il est déterminé par une multitude de critères et de déterminations personnelles, socioidentitaires et symboliques de la communauté en question et dont l'usage s'est étendu du web social Facebook à la sphère sociale du sujet utilisateur ;

4/Le chant des slogans contestataires du « hirak »

La présente contribution se propose d'interroger les chants contestataires propres au hirak algérien, mouvement de contestation populaire amorcé en février 2019. L'intérêt sera centré, ici, sur l'une des catégories du chant que j'appellerai le « chant-slogan ». Dotés d'un aspect organisateur du mouvement, ces chants sont actualisés pour servir d'autres desseins. Usant de stratégies linguistico-discursives plurielles, le chant-slogan remplirait, dès lors, des fonctionnalités diverses organisant le mouvement et renseignant sur son ingénierie. Cet article se consacrera précisément à l'exploration de l'une de ses fonctionnalités organisationnelles. S'agissant d'un discours persuasif dans la mesure où la lutte par le langage constitue une (ré) action dont la visée globale serait d'agir pour changer un état des choses, toute une rhétorique est déployée pour atteindre pacifiquement et progressivement les objectifs escomptés.

La présente étude est à situer dans le champ de l'analyse du discours. À travers l'analyse des actes de langage (Austin, 1962) constitutifs de ces chants et la détermination de leurs visées pragmatiques, on pourrait tenter d'identifier l'auditoire visé et mieux interpréter les fonctionnalités organisationnelles du chant-slogan dans ce contexte précis de mobilisation politique. qui serait la (les) cible(s) de ces chants-slogans et surtout quelles sont les fonctionnalités organisationnelles qu'il remplit?L'étude s'appuie sur un corpus des données observables constituées à partir de la plateforme « You tube». Il est constitué d'une série de chants-slogan produits lors des manifestations hebdomadaires du hirak ayant lieu dans l'espace public algérois. Pour explorer les logiques internes de cette production langagière particulière et complexe, J'examinerai, à partir d'un corpus puisé du contexte numérique, la production et la circulation des chants-slogans dans l'espace public algérois.

5/Discours épilinguistiques autour de la promotion de l'Anglais au profit du Français en Algérie, à travers les commentaires Facebook.

Résumé : Dans la lignée de mes travaux précédents sur l'usage des langues en Algérie, cette contribution se limitera à un examen des changements de statut des langues qui sont d'abord orientés dans le sens de substitution, d'un passage progressif vers un enseignement des filières techniques et scientifiques en Anglais, employé depuis Septembre 2024, par l'ensemble de la communauté universitaire comme langue d'enseignement. Il s'agit ensuite, de présenter ce panorama à partir de l'étude des représentations que se font les Algériens des deux langues en

présence en l'occurrence le Français et l'Anglais à travers une analyse des discours épilinguistiques recueillis sur une page Facebook « Le cercle des bordjiens et leurs amis » comptant 16 628 Membres, du 29/02/2024 au 05 /03 /2024. Cet examen donne une vue d'ensemble sur les perceptions, les attitudes valorisantes des locuteurs algériens vis-à-vis du Français en dépit de cette nouvelle mesure du ministère de l'enseignement supérieur. L'étude menée met également en évidence le rôle des représentations linguistiques dans l'usage des deux langues en question mais qui ne sont pas d'emblée déterminantes dans la promotion de l'Anglais, ni dans le processus de sa réhabilitation.

6/l'impact du contexte discursif sur l'usage de la métaphore dans le discours journalistique

La presse écrite en Algérie représente un discours riche de stratégies qui servent à transmettre l'information ; Ces dernières sont guidées par des conditions sociales. Nous avons tenté d'étudier l'environnement discursif d'une seule stratégie, qui est la métaphore. En essayant d'élucider l'influence de la société sur l'usage de la métaphore mais aussi l'influence de l'emploi de la métaphore dans certaines conditions sociales.

CORRIGE TYPE

a) Objectif de chaque recherche

1. **Raï/Rap et alternance codique**→ Expliquer les raisons du mélange arabe/français (emprunt et alternance) dans les chansons.
2. **Ironie dans le discours journalistique**→ Montrer le rôle argumentatif de l'ironie dans la chronique journalistique.
3. **Pseudonymes sur le web**→ Identifier les motivations et les enjeux sociolinguistiques du choix des pseudonymes.
4. **Chants-slogans du Hirak**→ Analyser les fonctions discursives et organisationnelles des chants contestataires.
5. **Discours sur l'anglais vs français**→ Étudier les représentations linguistiques à travers les commentaires Facebook.
6. **Métaphore dans la presse**→ Examiner le du contexte social sur l'usage de la métaphore-discours journalistique

b) Niveaux de recherche

1. → Sociolinguistique + analyse du discours
2. → Analyse du discours (argumentation)
3. → Sociolinguistique + onomastique
4. → Analyse du discours + pragmatique
5. → Sociolinguistique (représentations linguistiques)
6. → Analyse du discours (rhétorique)

c) Recherche la plus intéressante : Toute réponse argumentée est acceptable.

TD 03 : LES HYPOTHESES DE RECHERCHE

Les hypothèses sont des points de départ pour votre recherche. Elles vous permettront de construire un cadre théorique solide. Formulez deux hypothèses pour chaque thématique proposée dans les trois activités du « TD 1 »,

Etude morphosyntaxique des écrits des panneaux publicitaires de la ville

- **Hypothèses:**

- Les panneaux publicitaires utilisent principalement des phrases courtes, des impératifs et des constructions nominales pour capter rapidement l'attention du public.
- L'emploi de néologismes et d'anglicismes est fréquent dans les panneaux publicitaires pour créer un effet de modernité et de mondialisation.
- Les panneaux publicitaires privilégient un registre de langue simple et direct, adapté à un public large et hétérogène.

2. Etude de la structure narrative de quelques contes populaires de la région de l'Est algérien

- **Hypothèses:**

- Les contes populaires de l'Est algérien présentent une structure narrative similaire à celle des contes traditionnels d'autres cultures (introduction, développement, conclusion).
- Les personnages des contes sont souvent des archétypes (héros, méchants, sages) qui incarnent des valeurs culturelles spécifiques.
- Les contes populaires véhiculent des enseignements moraux et des valeurs sociales transmises de génération en génération.

3. Le statut de la langue française, face à l'Anglais, dans les filières scientifiques à l'Université algérienne

- **Hypothèses:**

- L'anglais est en train de supplanter progressivement le français comme langue de référence dans les publications scientifiques et les échanges académiques.
- Les étudiants algériens éprouvent des difficultés à concilier l'apprentissage des deux langues, ce qui peut avoir un impact sur leurs performances académiques.
- Le choix de la langue d'enseignement dépend en partie de la discipline scientifique et du niveau d'étude.

4. Lecture sociolinguistique de la trame romanesque de Malika Mokeddem

- **Hypothèses:**

- Les romans de Malika Mokeddem mettent en scène les tensions identitaires et les enjeux de la condition féminine dans la société algérienne.
- L'auteure utilise un langage riche et imagé pour décrire les réalités sociales et culturelles de son pays.
- Les dialogues dans les romans de Malika Mokeddem reflètent les variations linguistiques liées aux différentes classes sociales et aux contextes d'interaction.

5. Etude comparative des proverbes et dictons français et algériens sur le thème de « la femme »

- **Hypothèses**

- Les proverbes et dictons sur la femme reflètent les stéréotypes et les rôles traditionnellement attribués aux femmes dans chaque culture.
- Les proverbes algériens sur la femme mettent davantage l'accent sur les valeurs familiales et la soumission à l'autorité masculine que les proverbes français.
- Les proverbes sur la femme ont évolué au fil du temps en fonction des changements sociaux

1.**Etude sociolinguistique des affiches publicitaires dans les rues de la ville d'ORAN**

- Hypothèse 1 : Les affiches publicitaires dans les rues d'Oran reflètent la diversité linguistique de la ville, avec une prédominance de l'arabe et du français.
- Hypothèse 2 : Les affiches publicitaires dans les rues d'Oran ciblent principalement les jeunes adultes, avec des messages qui jouent sur les émotions et les valeurs de la modernité.
- Hypothèse 3 : Les affiches publicitaires dans les rues d'Oran révèlent les tensions entre la culture traditionnelle et la culture moderne, avec des références à la religion et à la identité nationale.

2. Analyse discursive des titres journalistiques dans la presse francophone écrite pendant le Hirak : Cas du journal Libération

- Hypothèse 1 : Les titres journalistiques du journal Libération pendant le Hirak ont utilisé une rhétorique de la résistance et de la lutte pour dénoncer la répression et soutenir les revendications du peuple.
- Hypothèse 2 : Les titres journalistiques du journal Libération pendant le Hirak ont privilégié une approche émotionnelle pour susciter la compassion et la solidarité envers les manifestants.
- Hypothèse 3 : Les titres journalistiques du journal Libération pendant le Hirak ont révélé des tensions entre la presse écrite et les réseaux sociaux, avec des références à la désinformation et à la manipulation médiatique.
- **3. Etude onomastique des toponymes de la région de Zemmoura**
- Hypothèse 1 : Les toponymes de la région de Zemmoura révèlent une histoire de colonisation et de résistance, avec des noms de lieux qui évoquent la présence française et la lutte pour l'indépendance.

Hypothèse 2 : Les toponymes de la région de Zemmoura reflètent la diversité culturelle et linguistique de la région, avec des noms de lieux qui proviennent de l'arabe, du berbère et du français.

Hypothèse 3 : Les toponymes de la région de Zemmoura ont évolué au fil du temps, avec des changements de noms de lieux qui reflètent les transformations politiques et sociales de la région.

4. L'emprunt lexical dans le domaine commercial : cas de la publicité

- Hypothèse 1 : Les emprunts lexicaux dans le domaine commercial sont principalement issus de l'anglais, et reflètent la globalisation économique et la domination de la langue anglaise dans le commerce international.

- Hypothèse 2 : Les emprunts lexicaux dans le domaine commercial sont souvent accompagnés de créations lexicales nouvelles, qui répondent aux besoins spécifiques du marché et de la publicité.
- Hypothèse 3 : Les emprunts lexicaux dans le domaine commercial varient en fonction du produit ou du service commercialisé, avec des emprunts plus fréquents pour les produits de luxe et les services de pointe.

5. Les néologismes et la création lexicale pendant la période de Pandémie du COVID 19

- Hypothèse 1 : La pandémie du COVID 19 a entraîné une explosion de néologismes et de créations lexicales nouvelles, qui reflètent la nécessité de communiquer rapidement et efficacement sur les mesures de santé publique.
- Hypothèse 2 : Les néologismes et les créations lexicales pendant la pandémie du COVID 19 sont principalement issus de l'anglais, et ont été adoptés dans de nombreuses langues, y compris le français.
- Hypothèse 3 : Les néologismes et les créations lexicales pendant la pandémie du COVID 19 ont évolué au fil du temps, avec des changements de vocabulaire qui reflètent les mutations du virus et les adaptations des mesures de santé publique.

6. Les phénomènes de création lexicale chez les Youtubeurs algériens

- Hypothèse 1 : Les Youtubeurs algériens créent de nouveaux mots et expressions qui reflètent leur identité culturelle et leur expérience de la vie quotidienne en Algérie.
- Hypothèse 3 : Les phénomènes de création lexicale chez les Youtubeurs algériens varient en fonction du genre et du style de contenu, avec des créations lexicales plus fréquentes dans les vidéos de divertissement et de humour.

TD 04 : DEFINITION DU CORPUS EN « ANALYSE CONVERSATIONNELLE »

Texte support :

Autour de la notion de corpus

On admet généralement les définitions suivantes du terme corpus : « ensemble d'énoncés qu'on soumet à l'analyse et qui constitue le corpus de la recherche » ou encore « ensemble fini d'énoncés réels réuni en vue de l'étude d'un phénomène linguistique », Dubois (1994,123). Ces définitions ne mentionnent pas les problèmes et les choix qui président à la constitution du corpus. Qu'en est-il de la définition du corpus dans le domaine de l'analyse conversationnelle ? D'autres questions se lient à celle-ci : est-ce que, en analyse conversationnelle, le corpus s'impose au chercheur inscrit dans ce domaine ou est-ce la théorie ou encore la méthodologie qui lui impose les critères de son recueil ? S'agit-il réellement de données ou de fabrication de données ?

Ce qu'il faut savoir avant toute chose est que, le corpus dont il est question en analyse conversationnelle est un ensemble de données orales. Recueillir ou étudier des données orales n'est guère une tâche facile à accomplir, car c'est là où tous les paramètres qui concourent à sa réalisation et sa compréhension entrent en jeu : il s'agit de la langue orale elle-même, du contexte, de la personne ou des personnes qui parlent, de la culture, etc.

Le corpus est-il un ensemble de données ou une fabrication de données ? Quand on parle de données, on comprend que le corpus est donné, au chercheur qui veut bien le saisir et l'étudier. Toutefois, on parle aussi de fabrication de données, une expression qui laisse entendre que sur le terrain on n'a pas affaire à des données, mais à des réalisations langagières que le chercheur manipule et façonne selon les besoins de sa recherche. Dans ce sens, Mondada (1998, 86) signale que le corpus est « le résultat d'un long processus de « fabrication » plus que le résultat d'un simple « recueil » de « données » dont on fait souvent comme si elles étaient déjà là, offertes au chercheur qui veut bien s'en saisir (d'où les métaphores de l'enregistrement » ou de la « captation ») ».

Des travaux m'ont permis d'observer une évolution dans la « fabrication » du corpus. Outre les problèmes que j'ai rencontrés, pour le recueil lui-même, pour la transcription de mon corpus, je souhaite ici porter un regard critique sur le recueil du corpus et de montrer tous les profits que j'ai pu en tirer, en l'occurrence le choix des interactants et le choix de la situation d'enregistrement.

Les règles conversationnelles sont acquises progressivement dès le bas âge, mais elles ne sont pas l'objet d'un apprentissage scolaire c'est pourquoi « l'objectif de l'analyse conversationnelle est précisément d'expliciter ces règles en tous genres qui sous-tendent le fonctionnement des échanges communicatifs ; ou, en d'autres termes, de déchiffrer la « partition invisible » qui guide (tout en leur laissant une large marge d'improvisation) le comportement de ceux qui se trouvent engagés dans cette activité polyphonique complexe qui est la conduite d'une conversation ». (Kerbrat-Oreccioni Catherine 1998, 9). Il faut souligner que les règles conversationnelles varient selon les sociétés et les cultures, dépendent du contexte et sont de nature diverse. Par conséquent, le chercheur inscrit dans ce domaine doit prendre en considération tous ces paramètres dans son recueil de corpus.

En analyse conversationnelle, on insiste sur l'importance « d'observer les activités des participants dans des événements de la vie sociale ordinaire, dans des naturally occuring interactions, c'est-à-dire des interactions qui auraient eu lieu même en l'absence du chercheur et qui n'ont pas été élicitées ou orchestrées par lui en vue de leur enregistrement ». (Mondada, 2005, 77). Dans cette perspective, il est clair que le chercheur doit prendre en considération cela, car il n'a pas le droit de bouleverser les habitudes des personnes en vue d'un enregistrement, il est question du culturel et surtout du naturel. C'est dans ce sens que Mondada, invite à travailler sur « des enregistrements – audio ou vidéo – d'interactions sociales, c'est-à-dire sur des données permettant de documenter l'émergence et le déploiement de ces pratiques dans le temps – et donc ni sur des descriptions de celles-ci (dans des entretiens, dans des notes prises par le chercheur) ni sur des produits de celles-ci (par exemple dans des textes issus d'une manière ou d'une autre de l'activité)… », **Mondada (2005 ,77).** »

QUESTIONS AUTOUR DU CORPUS

- Comment le texte définit-il le terme "corpus" en linguistique, et quelles distinctions sont faites entre les énoncés réels et les données fabriquées ?
- Quelles spécificités du corpus sont mises en avant dans le contexte de l'analyse conversationnelle ?
- Quelles sont les implications des choix méthodologiques sur la constitution du corpus en analyse conversationnelle ?
- Comment le texte décrit-il le rôle du chercheur dans la collecte et l'analyse des données du corpus ?
- Quelles sont les finalités de l'analyse conversationnelle telles que décrites dans le texte, et comment cela influence-t-il la sélection du corpus ?

CORRIGE TYPE

Définition du corpus : Dans le texte, on définit le corpus comme un ensemble d'énoncés soumis à l'analyse. Il fait une distinction entre les énoncés réels (données brutes) et les données fabriquées, soulignant que le corpus est souvent le résultat d'un processus de fabrication plutôt que d'un simple recueil de données.

Analyse conversationnelle : Dans l'analyse conversationnelle, le corpus est un ensemble de données orales qui nécessite une attention particulière aux contextes d'interaction et aux dynamiques sociales, en mettant l'accent sur la nature authentique des échanges.

Choix méthodologiques : Les choix méthodologiques influencent la constitution du corpus, car ils déterminent comment les données sont recueillies et analysées, notamment en ce qui concerne la sélection des interactants et des situations d'enregistrement.

Rôle du chercheur : Le chercheur est présenté comme un acteur qui façonne les données plutôt que comme un simple observateur. Il doit manipuler les données avec précaution pour ne pas altérer les interactions naturelles.

Règles conversationnelles : Il est crucial de prendre en compte les règles conversationnelles car elles guident le comportement des participants et influencent la dynamique des échanges, ce qui est essentiel pour une analyse précise.

Naturally occurring interactions : Ce concept désigne des interactions qui se produisent naturellement, sans intervention du chercheur. C'est important car cela permet d'étudier des échanges authentiques, reflétant des pratiques sociales réelles.

Le texte mentionne une évolution dans la fabrication du corpus, soulignant les défis liés à la collecte et à la transcription des données, qui peuvent affecter la représentation des interactions.

Le contexte culturel et les spécificités des interactants influencent la constitution du corpus en façonnant les règles conversationnelles et les attentes sociales, ce qui doit être pris en compte lors de l'analyse.

Transcription des données : Les enjeux liés à la transcription incluent la fidélité à l'oralité et la nécessité de capturer les nuances des échanges, ce qui est essentiel pour une analyse rigoureuse.

Objectifs de la recherche : Les finalités de l'analyse conversationnelle visent à expliciter les règles sous-jacentes aux échanges communicatifs, influençant ainsi la sélection du corpus pour garantir la pertinence des données analysées.

TD 05 : CRITERES DE RECUEIL DE CORPUS

ACTIVITE 01

Quels sont les objectifs possibles pour lesquels un corpus peut être constitué ? Citez-en au moins 3.

Quels sont les trois types de conditions qu'un corpus doit vérifier pour être pertinent ?

Quels outils ou méthodes pourraient faciliter l'analyse d'un corpus volumineux ?

ACTIVITE 02

1. Imaginez que vous devez constituer un corpus pour l'étude des néologismes dans les réseaux sociaux. Quels critères appliqueriez-vous pour assurer la pertinence, la représentativité et l'homogénéité ?
2. Vous travaillez sur un corpus d'entretiens pour analyser les expressions idiomatiques en français régional. Quels principes de régularité et d'exhaustivité suivriez-vous ?
3. Proposez une méthode d'échantillonnage pour un corpus visant à analyser les articles scientifiques publiés sur l'intelligence artificielle au cours des 10 dernières années. « analyse de Discours de presse sur l'intelligence artificielle »

CORRIGE TYPE

1. Objectifs d'un corpus (au moins 3) :

Décrire les usages réels de la langue.

Vérifier ou tester des hypothèses de recherche.

Analyser des phénomènes linguistiques (lexique, discours, variation).

2. Conditions de pertinence d'un corpus :

Représentativité (refléter la réalité étudiée)

Homogénéité (cohérence des données)

Pertinence/adéquation (adapté à l'objectif de recherche)

3. Outils/méthodes pour corpus volumineux :

Logiciels d'analyse (ex : AntConc, Sketch Engine)

Traitement automatique (TAL)

Méthodes statistiques et lexicométriques

ACTIVITÉ 02

1. Corpus sur les néologismes (réseaux sociaux) :

Pertinence : sélectionner des posts contenant des créations lexicales récentes

Représentativité : varier les plateformes, profils et périodes

Homogénéité : limiter à une langue/communauté ou type de discours précis

2. Corpus d'entretiens (expressions idiomatiques) :

Régularité : mêmes conditions d'enquête (type de questions, contexte)

Exhaustivité : recueillir un maximum d'occurrences sans sélection arbitraire

Transcription fidèle et systématique

3. Méthode d'échantillonnage (IA sur 10 ans) :

Échantillonnage chronologique (réparti sur 10 ans)

Échantillonnage stratifié (types de presse, disciplines, pays)

Sélection d'un nombre représentatif d'articles d'actualité , selon mots-clés (ex : "intelligence artificielle")

TD 06 : L'OBSERVATION

ACTIVITE 01

A partir du modèle proposé « observation en situation », réalisez dans le cadre d'une enquête par « observation » en précisant les étapes :

- « Etude linguistique du parler des jeunes sur les réseaux sociaux » ou
- Etude des néologismes dans les pratiques langagières des jeunes

Grille d'observation pour une recherche en Master 2 :

- *Définir le lieu d'observation : Université, classe :2ème agronomie. Agriculture appliquée*
- Contexte d'étude Objectif : Etude de l'Effet de l'alternance codique (Anglais/français) sur la compréhension des cours scientifiques.
- Type d'observation : participante/directe.
- Matériel d'observation : Grille d'observation.
- Protocole d'observation :

Date /Durée de la séance /Matière scientifique /Nombre d'étudiants présents

A. Utilisation de l'alternance codique par l'enseignant

Critères	Oui/Non	Fréquence (faible, modérée, élevée)	Notes/Exemples
Passage de l'anglais au français pour clarifier un concept scientifique			
Utilisation de termes techniques en anglais suivis d'une explication en français			
Alternance codique pour gérer les interactions (questions/réponses, encouragements)			

B. Réaction des étudiants

Critères	Oui/Non	Fréquence (faible, modérée, élevée)	Notes/Exemples
Manifestations de compréhension lors de l'alternance codique (gestes, participation, expressions faciales)			
Augmentation de la participation active après une alternance codique			
Demandes explicites de clarification dans l'une ou l'autre langue			

C. Impact sur la compréhension des concepts scientifiques

Critères	Oui/Non	Fréquence (faible, modérée, élevée)	Notes/Exemples
Amélioration apparente de la compréhension grâce à l'explication bilingue			
Difficulté persistante malgré l'alternance codique			
Capacité des étudiants à reformuler les concepts dans l'une ou l'autre langue			

D. Autres éléments contextuels

Critères	Oui/Non	Fréquence (faible, modérée, élevée)	Notes/Exemples spécifiques
Lien entre le niveau de langue des étudiants et leur compréhension (anglais/français)			
Impact de l'alternance codique sur la durée de la séance			
Différences de réaction selon les groupes d'étudiants			

Corrigé type

Cadre général de l'observation	
Élément	**Description**
Lieu d'observation	Réseaux sociaux (Facebook, Instagram, TikTok)
Public ciblé	Jeunes (18–25 ans, étudiants universitaires)
Objectif	Étudier les néologismes utilisés par les jeunes et leurs fonctions
Type d'observation	Observation directe non participante
Matériel	Grille d'observation + captures d'écran
Période	à préciser :

2. Protocole d'observation

Identification des néologismes			
Critères	**Oui/Non**	**Fréquence**	**Notes / Exemples**
Apparition de mots nouveaux			
Créations hybrides (arabe/français/anglais)			
Détournement de mots existants			
Usage d'abréviations ou formes réduites			
Procédés de création lexicale			

Critères	**Oui/Non**	**Fréquence**	**Notes / Exemples**
Emprunts à l'anglais			
Dérivation / composition			
Verlan / transformations phonétiques			

Créations humoristiques ou expressives			

Fonctions des néologismes			
Critères	**Oui/Non**	**Fréquence**	**Notes / Exemples**
Fonction identitaire (marquer l'appartenance)			
Fonction ludique / humoristique			
Fonction expressive (émotions, opinions)			
Fonction économique (gain de temps)			

ACTIVITE 02 :

Précisez le cadre méthodologique et le corpus de cette étude :

Violence verbale sur facebook : -blasphèmes ,juron, insultes,

Notre intervention s'articule autour du phénomène de la violence verbale sur les réseaux sociaux plus précisément sur facebook. Ce dernier est l'un des plus grands réseaux virtuels au monde, il permet d'échanger, de communiquer, de partager : des opinions, des photos, des vidéos, des messages... mais parfois il devient aussi un espace de confrontation pollué par des critiques, des insultes, des invectives, des propos haineux, des attaques, des diffamations, des propos homophobes et racistes, etc., où se déchaîne aussi la violence. Dans cet article, nous analysons les expressions violentes utilisées par les kabyles dans leurs échanges. À partir de l'analyse d'un corpus constitué de commentaires des membres de la page « Les amis de la Kabylie », nous avons essayé de comprendre le fonctionnement du phénomène de la violence verbale. Cette analyse sera double : d'une part, nous mettrons l'accent à travers les commentaires recueillis sur ce phénomène de violence numérique, d'autre part, nous nous intéressons aux expressions kabyles les plus utilisées par les facebookers

ACTIVITE 03 : Même consigne

LE FRANÇAIS AU CONTACT DE L'ANGLAIS DANS le discours d'un blogueur/ youtubeur…

LE PSEUDONYME CHEZ LA COMMUNAUTÉ D'INTERNAUTES ALGÉRIENS SUR LE WEB SOCIAL : QUELS CHOIX ET QUELS ENJEUX SOCIOLINGUISTIQUES ?

l'usage du pseudonyme (forums, tchats, réseaux sociaux numériques)

PÉCIFICITÉS TECHNO-SÉMIO-LINGUISTIQUES DES COMMENTAIRES EN LIGNE

CORRIGE TYPE

1. Violence verbale sur Facebook

Cadre méthodologique

- Approche : sociolinguistique et analyse du discours
- Type d'étude : qualitative (avec possibilité de quantification des occurrences)
- Méthode : analyse de contenu / analyse discursive des commentaires
- Outils : catégorisation (insultes, jurons, blasphèmes…), grille d'analyse

Corpus

- Nature : commentaires Facebook
- Source : page « Les amis de la Kabylie »
- Langues : kabyle (principalement) + autres langues possibles
- Données : échanges écrits contenant des expressions violentes
- Critères : pertinence (violence verbale), homogénéité (même page), représentativité (plusieurs commentaires/utilisateurs)

2. Français au contact de l'anglais (blogueur/youtubeur)

Cadre méthodologique :

- Analyse du discours + sociolinguistique
- Étude qualitative (alternance codique, emprunts)
- Grille d'analyse linguistique

Corpus :

- Vidéos / publications d'un blogueur ou youtubeur
- Transcriptions des contenus
- Données bilingues (français/anglais)

3. Pseudonyme chez les internautes algériens

Cadre méthodologique :

- Sociolinguistique + onomastique
- Analyse socio-discursive

Corpus :

- Pseudonymes collectés (Facebook, forums, tchats)
- Profils et usages associés
- Données variées (formes linguistiques diverses)

4. Usage du pseudonyme (forums, tchats, réseaux sociaux)

Cadre méthodologique :

- Approche comparative (plateformes)
- Analyse sociolinguistique

Corpus :

- Pseudonymes issus de différents espaces numériques
- Données contextualisées (type de plateforme, interaction)

5. Spécificités techno-sémio-linguistiques des commentaires en ligne

Cadre méthodologique :

- Analyse du discours numérique
- Approche sémiolinguistique

Corpus :

- Commentaires en ligne (réseaux sociaux, sites web)
- Données multimodales (texte, emojis, hashtags)

TD 07 : L'OBSERVATION EN LINGUISTIQUE DE TERRAIN

ACTIVITE

Etude de texte, extrait de « LINGUISTIQUE DE TERRAIN » de Philippe Blanchet

« On ne peut étudier les hommes qu'en communiquant avec eux, ce qui suppose que l'on partage leur existence d'une manière durable ou passagère. L'observation participante consiste à participer réellement à la vie et aux activités des sujets observés[55]. »

Je range dans ce type d'enquête les investigations interprétatives que l'on réalise en tant que lecteur/récepteur (= « l'auditeur interprétant » de Gumperz) dans les textes, notamment littéraires.

Selon le degré de connaissance du terrain et d'insertion dans la communauté observée, les modalités de l'observation varient, progressant par paliers successifs vers une participation accrue et directe aux échanges.

Les avantages de ce type d'enquêtes sont nombreux et particulièrement « rentables » en termes scientifiques :

– il permet de réduire au maximum le fameux paradoxe de l'enquêteur (la présence de l'enquêteur modifie les pratiques qu'il souhaite observer), puisque l'enquêté n'est pas en situation d'enquête ;

– il permet conjointement d'intégrer à l'analyse et d'assumer les effets produits par le sujet-chercheur dans son interaction avec les enquêtés, de comparer les pratiques par rapport aux discours sur les pratiques ;

– il permet d'enquêter de l'intérieur de l'interaction langagière et/ou de la communauté linguistique étudiée(s) et donc d'observer des phénomènes habituellement cachés aux « étrangers ».

On a là une alternative précieuse à la terrible injonction paradoxale si fréquente de l'enquêteur à son informateur : « soyez spontané ! ».

Cette méthode participante suppose bien sûr que l'enquêteur appartienne à la communauté linguistique étudiée ou, au moins, y soit suffisamment intégré, adopté. L'une de mes étudiantes-doctorantes, Nathalie Tréhel, a appelé fort à propos « méthode tupéroneuse » celle par laquelle à partir de son réseau de relations locales pré-existant, elle se fait introduire de foyer en foyer où l'on suscite de « mini-réunions » café-gâteaux, afin d'enquêter sur les pratiques du parler local dans sa commune de Bretagne gallo. Dans un champ de recherche comme le nôtre, portant sur des pratiques linguistiques diglossiques souvent « tabou » et prises, avec les locuteurs, dans des enjeux psycho-sociaux lourds, on comprendra que l'observation participante occupe une place privilégiée[58]. On se rappelle que W. Labov avait dû recourir à l'intermédiaire de l'un de ses étudiants noirs pour étudier les pratiques de l'anglo-américain des Noirs de Harlem[59].

La plupart des sociolinguistes travaillent sur des situations qu'ils vivent ou côtoient personnellement et la plupart des spécialistes de minorités linguistiques en sont issus ou y sont profondément intégrés, et pour cause ! Personnellement, je considère que l'ensemble de mes travaux sur les français régionaux, les alternances codiques, les pratiques linguistiques en contextes diglossiques et identitaires problématiques (Provence, Pays de Retz, Haute Bretagne, migrants) est de fait rendu possible, avec une certaine perspicacité, par le recours à cette observation participante. C'est du reste la méthode privilégiée par tous les auteurs de référence en ethnographie de la communication, interactionnisme et autre analyse conversationnelle, notamment en sociolinguistique par J. Gumperz.

À partir du moment où l'on veut étudier ce genre d'échanges, il est évident qu'il n'y a plus de sollicitation artificielle ni de méthode expérimentale suffisamment adaptée (ni, bien sûr, d'échantillonnage des informateurs a priori : on prend ce qui vient, on trie éventuellement après).

Enfin, en termes d'une éthique sociale et scientifique sur laquelle je reviendrai plus loin, le rôle d'acteur ici pris par « l'informateur », qui n'est pas un « cobaye » mais un réel producteur de savoir, et pour qui la recherche sur sa diglossie développe souvent une dynamique émancipatoire, me semble essentiel.

En revanche, l'observation participante présente certaines limites :

– la difficulté d'enregistrement des faits observés (la présence d'une caméra, d'un micro ou même

d'un simple calepin n'étant pas fréquente dans les situations de communication usuelles, notamment celles observées, et les enquêtes à appareil caché étant déontologiquement discutables) ; – l'implication personnelle de l'observateur, qui peut parfois induire une perception très orientée des phénomènes sans recul ni métaposition suffisante ultérieurement (à l'analyse) ; – l'aspect micro-sociolinguistique qui, certes riche de la complexité des paramètres observés finement, ne permet pas une représentativité suffisante du contexte sociolinguistique général ni la réunion d'un corpus ordonné permettant des techniques analytiques (par exemple en étude phonologique ou en statistiques).
C'est la raison pour laquelle sont nécessaires des procédures de distanciation, de comparaison et de contre-vérification par d'autres types d'enquêtes (semi-directives et directives) qui permettent des traitements plus quantitatifs. La distanciation et la comparaison sont également possibles grâce à une posture réflexive.

P. Blanchet, Linguistique de terrain, méthodes et théories2000, PP42-43

CONSIGNES

- Synthétisez le contenu du texte en 10 lignes(définir l'observation)
- Enumérez les avantages et les limites de l'observation

Résumé du texte

Dans cet extrait de Philippe Blanchet, l'auteur met en avant l'importance de l'observation participante en linguistique de terrain. Il souligne que l'étude des pratiques langagières nécessite une immersion du chercheur dans la communauté observée, afin de comprendre les usages réels de la langue dans leur contexte naturel. Cette méthode permet d'accéder à des interactions authentiques et de réduire l'effet perturbateur de la présence du chercheur. L'observation peut varier selon le degré d'intégration du chercheur, allant d'une simple présence à une participation active. Toutefois, cette approche exige une implication forte et une proximité avec les locuteurs. Enfin, l'auteur insiste sur la nécessité de compléter cette méthode par d'autres techniques d'enquête pour garantir la validité scientifique des résultats.

Avantages de l'observation participante	Limites de l'observation participante
- Réduction du paradoxe de l'enquêteur (moins d'influence sur les comportements).	• Difficulté d'enregistrement des données (contraintes techniques et éthiques).

- Accès à des données authentiques et naturelles. - Observation « de l'intérieur » des pratiques langagières. - Prise en compte de l'interaction entre chercheur et enquêtés. - Possibilité d'observer des phénomènes cachés aux observateurs externes.	• Implication subjective du chercheur (risque de biais). • Manque de recul analytique immédiat. • Faible représentativité (approche micro-sociolinguistique). • Difficulté à constituer un corpus structuré pour des analyses quantitatives.

TD 08 : L'ENQUETE PAR QUESTIONNAIRE

ACTIVITE 01 : élaborez un petit questionnaire sur les langues parlées ou pratiquées par un locuteur algérien .

CORRIGE TYPE

- Quelle est votre langue maternelle ?
- Quelles autres langues parlez-vous ou pratiquez-vous ? (Veuillez les énumérer)
- Pour chaque langue que vous parlez, veuillez indiquer votre niveau de compétence :
  - Débutant
  - Intermédiaire
  - Avancé
  - Bilingue
- Dans quels contextes utilisez-vous ces langues ?
- Travail
- Etudes
- Loisirs
- voyages, etc.)

-Avez-vous déjà rencontré des difficultés liées à la langue anglaise ? Si oui, lesquelles ?

- Comment avez-vous appris les langues que vous parlez ?

-l'école

- Autoformation

-Internet

Pensez-vous que l'anglicisation de l'enseignement scientifique à l'université est un avantage ? Pourquoi ?

ACTIVITE 02

Etude linguistique –Anthroponymique à base de questionnaire :

L'anthroponymie est une branche de l'onomastique qui s'intéresse à l'étude des noms de personnes. Proposez une démarche d'analyse et d'interprétation des prénoms de votre région par le biais d'un questionnaire ;

-Elaborez un questionnaire pour une catégorisation des motivations des choix des prénoms dans votre région.

CORRIGE TYPE

Ce questionnaire vise à explorer et comprendre les significations, les origines et les influences culturelles, religieuses, sociales qui façonnent nos choix de prénoms.

Votre prénom

- Quel est votre prénom ? Que signifie votre prénom ?
- Votre prénom est-il courant ou rare ?
- Votre prénom est :

- Moderne

-Archaique

-A connotation religieuse

Les motivations du choix de prénoms

Votre prénom est récurrent dans votre famille ? Pourquoi ?

- Comment les prénoms sont-ils choisis dans votre famille :

-tradition ancestrale

- signification

- Religion

- Aspect formel.

-Aspect moderne (La mode)

-Y a-t-il des prénoms que vous considérez comme typiques de votre région ? Lesquels ?

- Selon vous, quelles sont les principales motivations du choix de prénoms aujourd'hui :

mode,

religion,

culture,

autre

- Ya-t-il des prénoms traditionnels ou récurrents dans votre famille, région ? Si oui, lesquels ?

Les parents choisissent-ils des prénoms en fonction des prénoms déjà présents dans la famille ?

TD 09 : ENTRETIEN ET QUESTIONNAIRE : APPROCHE COMPARATIVE.

ACTIVITE : Etude de texte : *Enquêtes directives et semi directives « La linguistique du terrain de « P. Blanchet*

Entretiens : enquêtes semi-directives et directives[62]

Ces deux types d'enquête partagent la caractéristique fondamentale d'être organisés, présentés et réalisés auprès des informateurs en tant qu'enquêtes explicites. L'enquêteur travaille sur la base d'un questionnaire pré-établi, interroge l'informateur et recueille ouvertement les réponses, par enregistrement ou par écrit (de sa part comme parfois de celle de l'informateur).

La différence entre enquête semi-directive et enquête directive tient dans la formulation et la passation du questionnaire :

– l'enquête semi-directive est constituée de questions ouvertes auxquelles l'informateur peut répondre tout ce qu'il souhaite, lors d'un entretien, l'enquêteur se contentant de le suivre dans le dialogue (y compris si l'on s'écarte de la question pendant un certain temps) ;
– l'enquête directive est constituée de questions fermées dont les réponses sont prédéterminées et entre lesquelles, pour une question, l'informateur n'a qu'un choix limité : par exemple oui/partiellement/non – question fermée –, ou – question semi-fermée – « Comprenez-vous le provençal ? Pas du tout – un peu – à peu près – plutôt bien – très bien », ou encore gradation chiffrée à affecter à des réponses : « De quels parlers locaux celui/ceux du pays de Retz se rapproche(nt)-il(s) le plus ? Classez les réponses proposées par : 0 (pas du tout), 1 (très peu), 2 (un peu), 3 (moyennement), 4 (beaucoup), 5 (énormément) : – du Nord-Loire, – du Poitou vendéen, – du pays Nantais, – du pays de Clisson, – autres (précisez…) »

L'intérêt majeur en est évidemment le recueil d'informations attendues, estimées nécessaires à la compréhension du cas étudié, selon un cadre plus ou moins précis qui permet l'addition et le traitement statistique des données recueillies auprès de différents et parfois nombreux informateurs. On recueille des informations préalables sur les caractéristiques ethno-sociolinguistiques des informateurs (âge, origine, profession, langues connues, etc.), qui permettront d'établir d'éventuelles corrélations avec certaines réponses, ainsi que d'établir un échantillonnage statistique (y compris à partir d'un entretien semi-directif, car les informateurs proposent en général des réponses regroupables en types communs). Là, on doit choisir ses informateurs, sachant qu'il est important d'équilibrer variables et invariants (par exemple, si l'on fait varier la tranche d'âge, on garde la même localisation, etc.), et sachant qu'il est important de ne pas se limiter au type d'informateurs que le chercheur se représente a priori comme symptomatiques de ce qu'il recherche (sinon, il confirme artificiellement ses idées préconçues[63]). Enfin, l'entretien semi-directif permet également… »

Blanchet, Linguistique de terrain, méthodes et théories2000, P45

CORRIGE TYPE

Brève synthèse du texte

Le texte compare deux techniques d'enquête en linguistique : l'enquête **semi-directive** et l'enquête **directive (par questionnaire)**, en mettant en évidence leurs objectifs, leurs avantages et leurs limites.L'enquête **semi-directive** repose sur des questions ouvertes. Elle favorise l'expression libre de l'informateur, permet d'observer les pratiques langagières en situation et de mieux comprendre les représentations, les attitudes et les usages réels. Elle offre une richesse qualitative

L'enquête directive, fondée sur des questions fermées ou pré-cadrées, permet de recueillir des données comparables et quantifiables. Elle facilite l'analyse statistique et la vérification d'hypothèses. Toutefois, elle peut introduire des biais : les réponses peuvent être influencées par la formulation des questions, par les attentes supposées du chercheur ou par le contexte d'interaction.

Le texte insiste sur la nécessité de croiser les méthodes. L'approche interprétative privilégie l'observation et la contextualisation des pratiques, tandis que les enquêtes quantitatives permettent d'identifier des tendances générales. Aucune méthode n'est suffisante à elle seule : la validité de la recherche repose sur un équilibre entre objectivation, compréhension, contextualisation et vérification.

Enfin, le texte rappelle que toute enquête comporte des biais liés à l'interaction entre chercheur et informateur

Le rôle du chercheur est donc de les identifier, de les limiter et d'adopter une posture réflexive afin d'éviter de confirmer artificiellement ses hypothèses

Comparaison des deux techniques d'enquêtes en linguistique

La différence principale réside dans la formulation des questions : ouvertes dans l'entretien semi-directif, fermées ou graduées dans l'enquête directive.

Tableau 01 : comparaison entre enquête semi directive et directive

Enquête semi-directive	Enquête directive
Repose sur des questions ouvertes.	Repose sur des questions fermées ou semi-fermées.
L'informateur répond librement et peut développer.	Les réponses sont prédéterminées (oui/non, échelles chiffrées, choix multiples
L'enquêteur guide l'entretien mais suit le fil du discours.	L'enquêteur suit strictement un questionnaire structuré.
Donne des données riches, nuancées et qualitatives.	Produit des données comparables et quantifiables.
Permet d'explorer les représentations, attitudes, pratiques.	Facilite le traitement statistique et les comparaisons.
Risque : discours plus difficile à exploiter	Risque : rigidité

TD 10: OUTILS, TECHNIQUES DE COLLECTE DE DONNEES EN LINGUISTIQUE

ACTIVITE : ETTUDE DE TEXTE

« En sciences sociales, l'analyse du discours s'est développée dans un rapport critique avec l'analyse de contenu, « une technique de recherche pour la description objective, systématique et quantitative du contenu manifeste de la communication » (Bardin 1993 : 21 cité dans Charaudeau et Maingueneau 2002 : 39) reposant sur une catégorisation thématique des extraits textuels et leur traitement statistique. Conçue comme une méthodologie distincte, l'analyse du discours s'appuie sur des marques linguistiques pour établir son interprétation du sens. En outre, elle procède à une analyse des marques énonciatives attestant des actes de langage et de l'ancrage contextuel du discours.

Elle tient aussi compte des genres comme autant de pratiques sociales de communication prenant place dans des dispositifs énonciatifs ainsi que de l'interdiscursivité. Toutefois, avec la diversification des courants linguistiques et l'ouverture de l'analyse de contenu à des techniques s'en inspirant, de plus en plus d'études marient analyses de contenu et du discours (Charaudeau et Maingueneau 2002 : 40). Dans les années 1960, l'analyse du discours s'est dégagée de la linguistique structurale en suivant trois directions. D'une part, Harris, disciple du linguiste Bloomfield, a systématisé les procédures d'analyse structurale héritées de la description formelle des langues amérindiennes par l'anthropologie. Sa méthode consistait à décrire des régularités dans la distribution de classes d'équivalence, soit des mots ou séquences de mots survenant dans des environnements textuels équivalents. S'en inspirant, la méthode harrissienne en analyse du discours consistait à construire un corpus de phrases autour de termes pivots – c'est-à-dire des termes jugés représentatifs d'une formation discursive – et, à travers une série de transformations visant à réduire la diversité syntaxique, à comparer leur emploi dans des formations discursives concurrentes. De même, l'analyse automatique du discours de Pêcheux cherchait à réduire les discours en énoncés élémentaires afin de permettre un traitement informatique (Charaudeau et Maingueneau 2002 : 378). D'autre part, le modèle actantiel de Greimas, qui généralisait les travaux de Propp sur les contes populaires russes qui influenceront l'anthropologie linguistique et la folk linguistics, consistait à dégager six actants (sujet, objet, destinateur, destinataire, adjuvant, opposant) et leurs relations selon les trois axes du désir, du savoir et du pouvoir afin d'analyser des énoncés narratifs et, plus largement, des récits. Ces deux méthodes cherchaient à décrire des régularités distributionnelles ou sémantiques échappant au cadre de la phrase (Sarfati 2005 : 12). Enfin, dans le monde anglo-saxon, l'analyse du discours s'est détachée de la linguistique structurale sous l'influence, d'un côté, de l'ethnométhodologie et de l'ethnographie de la communication – courants conçus dans les années 1960 respectivement par Garfinkel ainsi que par Gumperz et Hymes qui inspireront l'analyse conversationnelle de Sacks et ses collaborateurs dans les années 1970 – et de l'autre, de la sociolinguistique issue des travaux de Labov (Charaudeau et Maingueneau 2002 : 233, 236-237; Quinn 2005).

- Elaborez un tableau comparatif mettant en évidence les différences et les similitudes entre l'analyse de contenu, la technique par questionnaire et l'entrevue.

- Choisissez un court extrait d'une recherche empririque en AD (par exemple, un article de journal, un discours politique, discours numérique…etc) :- Appliquez la méthode harrissienne pour identifier les termes pivots et analyser leur fonctions , leurs distribution.

-Appliquez le modèle actantiel de Greimas à un récit simple (par exemple, un conte de fées).

-Analyse **des Marques Énonciatives** :

- Expliquez ce que sont les marques énonciatives et leur importance dans l'analyse du discours.
- Donnez des exemples de marques énonciatives et comment elles peuvent révéler l'ancrage contextuel d'un discours.

Elaborez un tableau comparatif mettant en évidence les différences et les similitudes entre l'analyse de contenu, la technique par questionnaire et l'entrevue

	Analyse de contenu	**Technique de l'Observation**	**3. Questionnaire**	**4/Entrevue**
Définition	Technique de recherche qui vise à décrire de manière systématique et objective le contenu de communications (textes, images, vidéos, etc.). Elle peut être quantitative (comptage de fréquences) ou qualitative (interprétation de thèmes).	Méthode de collecte de données qui consiste à observer et à enregistrer les comportements, les interactions et les phénomènes dans leur contexte naturel.	Instrument de collecte de données qui consiste en une série de questions posées (fermées, ouvertes, structurées) à un échantillon de personnes.	Méthode de collecte de données qui consiste en une conversation structurée ou non structurée entre le cherche ur et un

				participant.
Différences :	-Contrairement aux autres méthodes, elle se concentre sur des documents existants plutôt que sur des interactions directes avec des individus. -Elle permet d'analyser de grandes quantités de données de manière relativement efficace. -Elle peut révéler des tendances et des motifs qui ne sont pas immédiatement apparents.	Elle se concentre sur l'observation directe des comportements plutôt que sur les discours ou les opinions. o Elle peut être participante (le chercheur s'intègre au groupe observé) ou non participante (le chercheur observe de l'extérieur). o Elle permet de recueillir des données sur des comportements non verbaux et des interactions sociales.	-Il permet de recueillir des données standardisées auprès d'un grand nombre de personnes. -Il est généralement utilisé pour recueillir des données quantitatives, mais il peut également inclure des questions ouvertes pour recueillir des données qualitatives. -Il est moins flexible que l'entrevue, car les questions sont prédéfinies.	-Elle permet de recueillir des données approfondies et nuancées sur les expériences, les opinions et les perspectives des individus. -Elle est plus flexible que le questionnaire, car le chercheur peut adapter les question

				s en fonction des réponses du participant. -Elle permet d'explorer des sujets complexes et sensibles.
Similitudes	Comme l'observation et l'entrevue, elle vise à recueillir des données pour comprendre un phénomène social.	-Comme le questionnaire, elle peut être utilisée pour recueillir des données quantitatives (par exemple, le nombre d'occurrences d'un phénomène).	-Comme l'entrevue, il vise à recueillir des informations sur les opinions, les attitudes et les comportements des individus. -Comme l'analyse de contenu, il peut être utilisé pour identifier des tendances et	-Comme l'observation, elle permet de recueillir des données qualitatives sur les interactions sociales.

			des motifs dans les données.	-Comme le questionnaire, elle vise à recueillir des informations sur les opinions, les perceptions, les attitudes et les comportements des individus.

TD 11 : ANALYSE ET INTERPRETATION DES DONNEES DE CORPUS

ACTIVITE : ETUDE DE TEXTE

Traitement des donnees : analyse, comparaison et synthèse interprétative

Une fois le recueil des données effectué, en allant de la prise globale d'informations et d'indices de micro-situations de l'échantillon constitués par enquêtes directives, par observation participante, de questionnaire, arrive la phase la plus fine de la démarche, **l'analyse de ces données** ouvrant aussitôt sur l'élargissement que va produire la synthèse interprétative . Cette étape analytique fonctionne de faits ou chaque donnée est examiné en détail, trié, rangé dans un ensemble descriptif dont telle forme est une pièce. Cependant, et c'est là notamment qu'une approche systémique complexe interprétative diffère d'une approche analytique structurale classique, le travail du chercheur ne s'arrête pas là.

Du point de vue méthodologique, en effet, en cohérence avec le cadre théorique, le fonctionnement global de l'ensemble agit sur les parties. C'est le contexte global de la ou des situations de communication ;

Les procédures d'analyse mises en point depuis de nombreuses années dans les champs linguistique, sociologiques, anthropologiques…etc sont connues et établies. Il s'agit des procédures phonologiques, morphologiques, syntaxiques, etc. Des méthodes d'échantillonnage et de calcul statistique d'investigation structurale des langues notamment en linguistique. En matière d'analyse linguistique, on y recherche ce qui varie au moins autant que ce qui est stable, que ce soit dans une structure donnée (**il n'ya par exemple opposition phonologique que s'il y'a variation signifiante possible).** Ou dans les pratiques linguistiques en général (attention portée aux changements diachroniques ou aux diverses variations /variétés non standard d'une langue) , une approche phonologique, combinée à une approche éthno-sociolinguistique.

En termes d'analyse de discours, on préfère bien sur les techniques de contextualisation (rapport à l'ensemble d'un discours et à d'autres discours, ainsi qu'aux paramètres sociaux » aux décompositions sémantiques et structuralistes et aux analyses limitées à la phrase ou au mot.

Les techniques d'analyse des co-occurrences, des contraintes et stratégies pragmatiques des positionnements socio-institutionnels de l'énonciation d'une part , les analyses rapportant à une interaction, et pas seulement le discours et contexte idéologique et culturel, d'autres part , une analyse conversationnelle façon Gumperz et hymes, ludi, PY…etc Enfin, les chercheurs en linguistique ont régulièrement recours pour leurs travaux aux outils de description des mimiques, gestes, des postures, relevant du non-verbal accompagnant souvent l'expression verbale.

Une analyse quantitative à l'aide de chiffres joue un rôle mineur dans l'ensemble de cette méthodologie essentiellement qualitative, ou ils ne servent que de compléments d'enquêtes. Les spécialistes des analyses quantitatives conviennent eux-mêmes du fait que « au-delà du vocabulaire, la prise en compte du contexte peut dans la mesure ou c'est possible changer complètement la perspective et permettre une analyse de toute une autre ampleur. L'analyse textuelle devient alors en elle-même une blanche de l'analyse des données » .

Les procédures comparatives contribuent à réduire l'implication subjective du chercheur et offre des résultats satisfaisants en termes de compréhension d'un cas précis qu'en terme de généralisation c'est-à-dire **la synthèse, conçue** comme « la saisie globale des paramètres en jeu dans une perspective systémique complexe. Il ne s'agit pas d'un résumé final ,mais des procédures analytiques comparatives et interprétatives ». **L'interprétation,** dans notre domaine, fonctionne à deux niveaux. Au premier niveau, elle désigne l'objet de la sociolinguistique interactionnelle et ethnographie de la communication de Gumperz et **hymes « *il s'agir de découvrir les liens étudiés entre la perception des signes linguistiques de surface et l'interprétation d'expliquer la capacité humaine à contextualiser l'interprétation*…de mettre l'accent sur les stratégies qui gouvernent l'emploi par l'acteur des connaissances lexicales, grammaticales, sociolinguistiques et autres dans la production et l'interprétation des messages en contexte ».** Ce sont les interprétations par les

interlocuteurs des signaux ethno-sociolinguistiques (variation) comme indices signifiants dans une interaction langagière. Cette synthèse interprétative se réalise en :
-mise en relation de données langagières observées « objective » avec le contexte perçu de leur productions et avec d'autre part les données « subjectives » recueillies auprès des locuteurs.
-Recherche de récursivité des phénomènes et des fonctionnements. Analyse des relations entre variables
- Comparaison des résultats observés avec ceux attendus et interprétation des écarts : Comparaison des parties et du tout, selon données jusqu' à structuration. confronter les relations avec celles prévues dans les hypothèses.

Synthèse

La phase de traitement des données dans la recherche linguistique

1. Du recueil des données à l'analyse

Après la phase de collecte des données, le chercheur entre dans une étape essentielle : le traitement et l'analyse du corpus.

Les données peuvent provenir de différentes méthodes :

- enquêtes directives
- observation participante
- questionnaires
- enregistrements de situations de communication

Ces données constituent un échantillon représentatif de pratiques linguistiques qui sera ensuite analysé.

2. L'analyse comme étape centrale

L'analyse consiste à :

- examiner chaque donnée en détail
- classer et organiser les informations recueillies
- identifier les régularités et les variations linguistiques

Cette étape permet de transformer les données brutes en données interprétables scientifiquement.

II. Les méthodes d'analyse en linguistique de corpus

Les sciences du langage ont développé plusieurs procédures d'analyse linguistique, notamment :

1. L'analyse linguistique structurale

Elle repose sur l'étude des différents niveaux de la langue :

- analyse phonologique

- analyse morphologique
- analyse syntaxique
- analyse lexicale

Ces méthodes permettent d'identifier les structures et les variations linguistiques.

2. L'étude de la variation linguistique

La linguistique ne s'intéresse pas uniquement aux structures stables mais aussi aux variations

- variations phonologiques
- variations sociolinguistiques
- changements diachroniques
- variétés non standard d'une langue

Ces analyses permettent de comprendre l'évolution et la diversité des pratiques linguistiques.

3.L'analyse du discours et la contextualisation

Dans les approches contemporaines, l'analyse linguistique ne se limite pas à la phrase.

Elle privilégie :

- l'étude du discours dans son contexte
- les relations entre discours et société
- les interactions entre locuteurs

Les chercheurs utilisent notamment :

- l'analyse des co-occurrences lexicales
- l'étude des stratégies pragmatiques
- l'analyse des positionnements socio-institutionnels

4. L'analyse interactionnelle et ethnographique

Certains travaux s'appuient sur les approches de :

- John Gumperz
- Dell Hymes

Ces approches relèvent de :

- la sociolinguistique interactionnelle
- l'ethnographie de la communication

Elles visent à comprendre :

- comment les locuteurs interprètent les signes linguistiques
- comment le contexte social influence l'interprétation du discours

L'analyse peut aussi inclure :gestes ; mimiques ; postures

Ces éléments relèvent de la communication non verbale.

5. Analyse qualitative et analyse quantitative

Dans la linguistique de corpus :

1. L'analyse qualitative

Elle occupe une place centrale et consiste à :interpréter les données linguistiques, analyser les pratiques discursives, et comprendre les relations entre langue et société

2. L'analyse quantitative : repose sur :le calcul statistique ; la fréquence des formes linguistiques, l'analyse lexicométrique

Cependant, les données chiffrées restent souvent complémentaires à l'analyse qualitative.

LA SYNTHESE INTERPRETATIVE

La synthèse ne correspond pas simplement à un résumé des résultats.

Elle consiste à :

- établir des relations entre les données
- analyser les interactions entre variables
- interpréter les écarts entre résultats observés et hypothèses initiales

Elle repose sur une approche systémique, où le sens des données est compris en relation avec l'ensemble du contexte.

SYNTHESE

Dans la linguistique de corpus, le traitement des données constitue une étape essentielle de la recherche. Après la collecte du corpus, le chercheur procède à une analyse détaillée des données linguistiques en mobilisant différentes méthodes issues de la linguistique, de la sociolinguistique et de l'analyse du discours.

Cette analyse combine souvent des approches qualitatives et quantitatives, mais elle accorde une place centrale à la contextualisation des pratiques langagières. L'objectif n'est pas seulement de décrire les structures linguistiques, mais aussi de comprendre les variations, les interactions et les stratégies discursives des locuteurs.

La recherche aboutit finalement à une **synthèse interprétative**, qui consiste à mettre en relation les données linguistiques, les contextes sociaux et les hypothèses de départ afin de produire une compréhension globale des phénomènes étudiés.

TD 12: ETUDE DE CORPUS VARIES « EXTRAITS DE L'ASJP »

ACTVITE

ETUDE DE CORPUS URBAIN

Notre travail de recherche s'inscrit dans le cadre de la sociolinguistique urbaine. Il est centré d'une part sur la situation sociolinguistique d'une ville : Batna, et d'autre part sur les rapports entre les pratiques linguistiques des locuteurs et les contraintes de l'espace urbain. il s'agit d'une étude des influences du contexte urbain et plurilingue sur les différentes manifestations linguistiques. En effet, nous décrivons une situation plurilingue dans un espace urbain, avec des interactions complexes entre langues (français, arabe dialectal, chaoui), groupes sociaux et espace ; car comme le souligne Thierry Bulot, l'une des tâches de la sociolinguistique urbaine est de « *questionner les rapports entre langues (…) sous l'angle des contraintes du terrain urbain* » (2004,10). Pour réaliser ce travail de recherche sur le terrain, nous avons mené une enquête qui avait comme objectif de recueillir à partir d'un large échantillon de locuteurs les différentes opinions sur l'emploi des langues (français, chaoui, arabe dialectal) dans des situations réelles et de vérifier par la suite les productions linguistiques authentiques de ces mêmes locuteurs et surtout de mettre en valeur cette relation qui peut exister entre les pratiques linguistiques des locuteurs batnéens et leurs représentations spatiales et sociolinguistiques. Nous pouvons distinguer dans la collecte des données (corpus) trois niveaux d'information : la pré-enquête, (l'enregistrement) et le (le questionnaire). Le questionnaire et l'enregistrement sont des outils indispensables dans notre travail de recherche centré sur l'urbanité langagière.

Tout d'abord le discours enregistré des locuteurs batnéens. Ce discours a été enregistré dans différents quartiers de la ville de Batna et avec différents locuteurs (un groupe d'étudiants, des commerçants, des intellectuels, des analphabètes…) (voir annexe1). A partir des enregistrements, nous avons tenté de décrire les différentes pratiques linguistiques des différents locuteurs dans les différents quartiers de la ville pour pouvoir comprendre comment s'articulent les différents espaces urbains au sein de cette ville et aussi d'expliquer l'influence des représentations spatiales sur ces même pratiques. Sans oublier aussi qu'à travers ces mêmes enregistrements nous avons essayé de comprendre les attitudes et les perceptions des locuteurs batnéens vis-à-vis des langues, ainsi que leurs représentations spatiales. ce questionnaire a visé à dégager les représentations que les locuteurs batnéens se font de leur ville, des échanges linguistiques qui s'y déroulent. Les informations recueillies servent à compléter les enregistrements et permettent de mieux comprendre, les pratiques linguistiques de ces locuteurs et leurs stratégies adoptées dans la communication courante. La pré-enquête a consisté en plusieurs visites au sein de la communauté ciblée (la ville de Batna), visites au cours desquelles nous avons pu mener diverses activités avec les membres du groupe d'enquête : très souvent, nous nous rendons dans les différents magasins et sur les marchés de la ville de Batna pour faire soit des achats soit pour entretenir des discussions avec les gérants de ces magasins et même avec les clients. Cette démarche nous a permis tout d'abord d'approcher nos informateurs et par conséquence la délimitation de notre communauté du langage. C'est ainsi que nous avons pu identifier le groupe chaouiphone de celui des arabophones. Sans cette étape, la constitution d'un échantillon pertinent adapté à nos hypothèses aurait était impossible.

Annexe 1
Situation informelle
Enregistrement : II (Extrait) Lieu : Un magasin
Nombre de locuteurs : 4 jeunes filles.
Durée : Demi-heure
Date : 09/04/06

Thèmes de discussion : 1/ La cause berbère
2/ Activités préférées
A : " Fi waqt Boumedienne laqbayal hawlou bach imposiw alougha antaâhoum laqbaylia **mais** Boumedienne qmaâ hadhik athawra makhalahoumche adahoume lalhabs makhalahoumche houriat aray maâbrouche aâla raïhoume kh'las maqbalhoumche alougha el amazighiat kalougha mahalia watania maâtarafche biha kalougha watania ma khalahache **par exemple** makhalalhoumche foursa mathalane fi asahafa almaktouba wala el masmouâ kima **télévision**

wa **radio** bach yahadrou kima b'ghaw bilaqbaylia antaâhoum iâbrou âla rayhoum âla chouâourhoum aâla **même** hata le masrah wala hata alfane makhalahoumchem yahadrou blakhor **sauf quelques chansons** ali khalahoume yadhahrouha fi asabâinat **ça fait** âdat hadhi laqbaylia walaw yafardhou r'wahoum basah **le pouvoir** a'souloutat makhalatalhoum'che atrigue yafardhou rwahoum wibaynou atourath antaâhoum… **silence** ".

Annexe 02 : Le questionnaire

Age:Sexe: Féminin , Masculin Region d'origine:
Pour les jeunes: Scolarité: Origine des parents :
Mère : Père :Si vous n'êtes pas originaire de Batna;pourquoi êtes-vous venu à cette ville?
Diplôme:
Profession ou métier:
Si vous n'êtes pas originaire de Batna;pourquoi êtes-vous venu à cette ville?
1/ En quel quartier habitez-vous?
2/ Quelle est la langue employée au sein de votre quartier?
-Par vous: -Par les autres habitants:
3/ Décrivez votre quartier ?
4/Quelle est la langue que vous utilisez chez vous(à la maison)?
5/ Pensez vous qu'au sein de votre quartier les gens parlent la même langue que vous? Oui Non dites pourquoi?
6/ la ville de Batna est-elle composée de plusieurs quartiers? Oui
Non
Si oui citez-les?
7/Quelle est la langue utilisée dans chaque quartier?
9/ Où trouve-t-on plus de gens qui parlent autre chose que le chaoui?
10/ Où trouve-t-on plus de gens qui parlent autre chose que l'arabe dialectal?
11/ Pensez-vous que les batnéens parlent chaoui? Oui
NonDites pourquoi?
14/ Où avez-vous appris à parler arabe ?
-A l'école:
-A la maison avec les parents:-Dans le quartier avec les copains et les copines:
15/ Où avez-vous appris à parler chaoui ?
-A l'école:-A la maison avec les parents:
-Dans le quartier avec les copains et les copines:
16/ Où avez-vous appris à parler français ?
-A l'école:-A la maison avec les parents:-Dans le quartier avec les copains et les copines: 17/ Vous considérez-vous:
D'identité arabe uniquement: -D'identité chaoui uniquement:
-Plus que arabe que chaoui: -Plus que chaoui qu'arabe: -Autres:

TD13: OUTILS DE COLLECTE ET D'ANALYSE DE DONNEES :L'ANALYSE DE CONTENU

Thématique : Actes De Langage et Discours Politique

Cette recherche se veut une analyse du discours politique de deux présidents français, à savoir Chirak et Sarkozy lors de leurs visites en Algérie. L'approche adoptée est l'approche pragmatique puisqu'il s'agit de l'étude des actes de langage indirecte. Nous allons donc observer les différentes stratégies discursives et énonciatives mises en place pour véhiculer des questions relatives à la guerre d'Algérie. Quelles sont les stratégies discursives des deux présidents français face aux destinataires Algériens? Comment se manifestent les présupposées, les allusions, les sous-entendus dans «un discours présidentiel »?

CONSIGNE

A partir de cette approche comparative, décrivez le corpus et cadre méthodologique de cette étude.

Définissez les phénomènes pragmatiques étudiés, cités dans le tableau ci-dessous

	Discours de CHIRAC	Discours de Sarkozy
L'implicite	*-Le premier implicite est historique qui relate les retrouvailles entre l'Algérie et la France comme dans l'extrait suivant :« C'est l'émotion des retrouvailles qui nous éteint ». Cet énoncé comporte des éléments linguistiques qui codent l'implicite. Ce passage indique qu'il y a eu une coupure ou une rupture entre les deux pays à un moment donné. Cette rupture n'était pas volontaire car Chirac manifeste dans son propos des émotions fortes. Ces éléments nous ramènent à penser à la période du terrorisme qui a brisé les relations entre les deux pays.* Le deuxième implicite est religieux, il renvoie à la tolérance. Comme dans l'extrait suivant : *« Et je voudrais dire combien nous partageons votre vision d'un Islam tolérant ».* Dans cet énoncé, Chirac se réfère à l'attitude générale du peuple algérien qui rejette l'Islam intégriste. Cet énoncé révèle implicitement, que l'Algérie refuse l'Islam intégriste, et qu'elle combat pour défendre la tolérance.	Parmi ces implicites, nous citons deux. Le premier implicite est historique : *« C'est mon devoir de président de la république de le dire, de part et d'autres, il y a eu des douleurs, il y a eu des souffrances, il y a eu de la peine »* Cet énoncé fait référence aux victimes de la guerre de l'indépendance de l'Algérie. Les victimes sont de deux côtés : les Algériens, mais aussi les Français ayant quitté de force l'Algérie, les Harki… Le deuxième implicite est politique : *« Ils étaient donc plus intelligents et plus courageux que nous ? ».* Cette interrogation nécessite une prise en compte du contenu de tout le paragraphe qui traite du projet de l'union pour la méditerranéen. Quand le président dit « ils », il veut dire nos aïeux qui ont réussi de faire cette union. L'implicite se trouve dans le « nous ». Dans ce cas précis, il s'agit des responsables des deux rives : l'Algérie et la France.
Les	Les présupposés dans le discours du Président Chirac	Nous avons choisi dans le discours du président Nicolas Sarkozy deux présupposés. Le premier est historique :

Présupposés	Nous avons trouvé des présupposés dans le discours du Président Chirac .Nous illustrons ce constat par deux exemples. Le premier supposé est politique : **« Notre dialogue politique s'est enrichi ».** Le choix de ce passage est motivé par l'utilisation du temps employé : le passé composé. Cela présuppose que les deux pays ont fait du chemin depuis l'indépendance, qu'ils continuent de fructifier leurs relations, et ce malgré le fardeau de l'histoire. Le deuxième présupposé est politico-économique **:« L'ouverture de l'Algérie sur le monde ».** Ce passage présuppose que l'Algérie est coupée du monde extérieur. Le président Chirac parle effectivement ici des années quatre-vingt-dix où l'Algérie est isolée du monde.	« **Il s'est agi pour nos deux vieux pays si longtemps ennemis de se tourner vers l'avenir ».** Ce présupposé nous fournis beaucoup d'éléments linguistiques pour le décoder. L'utilisation de « longtemps ennemis » ou « se tourner vers l'avenir » nous pousse à induire ceci : il est temps d'enterrer la hache de guerre de penser au futur commun. Le deuxième présupposé est politique : « qu'il reconnaisse au peuple d'Israël qui a tant souffert le droit de vivre libre ». Les indices « reconnaisse », « tant souffert », « le droit » et« vivre libre », nous renvoient à deux présupposés : le premier est qu'Israël occupe légitimement la terre en Palestine. Le deuxième est que cet état a le droit de se défendre en cas d'agression des pays limitrophes. - Les verbes performatifs sont lus dans les passages de Sarkozy : Je veux lancer un appel à l'Occident pour qu'il se dépouille de toute volonté de domination (Sarkozy; le verbe performatif « dépouiller » permet de poser que le monde occidental est dominant. La construction discursive de « NOUS », de « VOUS » et d'« EUX » commence avec l'étiquetage des acteurs sociaux, à savoir: Nous « Les Algériens et les Français » quand il s'agit à l'appel à la fraternité; Vous « Les Algériens »:
Les sous-entendus	Le premier sous-entendu est politico-historique : *« Vous le savez mieux que beaucoup d'autres, nous faisons face à des incertitudes, des difficultés, des défis ».* Selon le contexte de cet énoncé, le Président sous-entend ici que l'Algérie devrait comprendre l'instabilité dont le monde d'aujourd'hui. Le deuxième sous-entendu est politico-économique : *« Mais il reste beaucoup à construire et nous devons aller plus loin dans cette ambition ».* Le contexte de ce paragraphe révèle que le Président sous-entend que les pays méditerranéens	Les sous-entendus dans le discours du Président Sarkozy Nous avons constaté la présence de plusieurs sous-entendus dans le discours du Président Sarkozy. Nous citons deux exemples. Le premier sous-entendu est historico-politique : *« Je ne suis pas venu nier le passé ».* Cet énoncé montre que le Président est conscient du poids du passé et qu'il n'a pas l'intention de l'oublier. Mais, il souhaite vivement avancer et penser aussi à l'avenir commun. Le deuxième sous-entendu est religieux : *« C'est cela le message de l'Islam que vous devez porter en Algérie et ailleurs. ».* Le Président sous-entend que l'Islam authentique est ancré dans l'histoire, et que cet Islam est compatible avec la laïcité et la démocratie.
L'insinuation	**L'insinuation** **Selon kerbrat-Orecchioni : « comme étant en général un sous-entendu malveillant** *« Privilégions ce qui nous unit, par le cœur et par la raison ».* Dans cet énoncé, le Président insinue de mettre en avant les bonnes choses partagées ensemble sans penser aux souffrances qui détruisent les relations entre les deux pays. Il est important pour lui de penser aux intérêts collectifs**.**	Le premier est politique : *« Ne pas reconnaître à Israël le droit de la liberté est une injustice ».* Dans cet extrait, le Président Sarkozy insinue aux responsables algériens de changer le regard et de tempérer leurs discours à l'égard de l'Etat d'Israël. La deuxième insinuation est politico-économique : *« Il est devenu vital de plus force à ce qui nous unit et d'arrêter de parler de ce qui nous oppose ».*
L'allusion	**L'allusion une « référence implicite et oblique à un élément extérieur à l'univers de l'énoncé » (2007 :117).** L'allusion dans le discours du Président Chirac	**-Absence d'allusion**

	Nous avons remarqué la présence de plusieurs allusions dans le discours du Président Chirac. Deux exemples nous permettent de les illustrer. La première allusion est politique : *« car la guerre est toujours, c'est vrai, un constat d'échec, toujours un drame ».*Ce passage fait allusion aux éxpériences de la guerre qui ne règlent jamais un conflit sans faire un nombre considérable de morts, de veuves et d'orphelins. Comme c'était le cas dans le conflit entre la Serbie et la Bosnie. *-« Laissez-moi saluer le courage et la dignité des Algériens qui, dans les épreuves, ont su rester debout, fière et dans l'honneur ».* *-« Nous vivons aujourd'hui, je crois, un moment d'histoire…..»*	

Le corpus étudié se compose de deux discours présidentiels. Le premier discours est prononcé par le Président français Jacques Chirac, le lundi 3 mars 2003, à Alger, dans le palais des nations. C'est un discours de quatorze pages, soit cinquante-trois paragraphes. Ce discours se réfère à plusieurs domaines : histoire, politique, culture et enseignement, religion et économie. Quant au deuxième discours est prononcé par le Président français Nicolas Sarkozy, le mercredi 5 décembre 2007, à Constantine, en Algérie, dans l'université de Mentouri. Ce discours contient sept pages, soit trente-huit paragraphes. Il a évoqué presque les mêmes domaines cités dans le discours du Président Chirac. L'analyse de notre corpus, est une « analyse de contenu des discours présidentiels», s'appuie sur cinq catégories de l'acte du langage indirect à savoir : les implicites, les présupposés, les sous-entendus, les insinuations et les allusions.

Analyse et interprétation du corpus

TD 14: TELLE THEMATIQUE, TEL CORPUS : Études de Cas.

CONSIGNE

Comment la méthodologie adoptée dans cette recherche permet-elle d'établir un lien entre les pratiques linguistiques des locuteurs batnéens et les contraintes de l'espace urbain dans un contexte plurilingue ?

CORPUS URBAIN

Notre travail de recherche s'inscrit dans le cadre de la sociolinguistique urbaine. Il est centré d'une part sur la situation sociolinguistique d'une ville : Batna, et d'autre part sur les rapports entre les pratiques linguistiques des locuteurs et les contraintes de l'espace urbain. il s'agit d'une étude des influences du contexte urbain et plurilingue sur les différentes manifestations linguistiques. En effet, nous décrivons une situation plurilingue dans un espace urbain, avec des interactions complexes entre langues (français, arabe dialectal, chaoui), groupes sociaux et espace ; car comme le souligne Thierry Bulot, l'une des tâches de la sociolinguistique urbaine est de « questionner les rap

ports entre langues (…) sous l'angle des contraintes du terrain urbain » (2004,10). Pour réaliser ce travail de recherche sur le terrain, nous avons mené une enquête qui avait comme objectif de recueillir à partir d'un large échantillon de locuteurs les différentes opinions sur l'emploi des langues (français, chaoui, arabe dialectal) dans des situations réelles et de vérifier par la suite les productions linguistiques authentiques de ces mêmes locuteurs et surtout de mettre en valeur cette relation qui peut exister entre les pratiques linguistiques des locuteurs batnéens et leurs représentations spatiales et sociolinguistiques.

Nous pouvons distinguer dans la collecte des données (corpus) trois niveaux d'information : la pré-enquête, (l'enregistrement) et le (le questionnaire). Le questionnaire et l'enregistrement sont des outils indispensables dans notre travail de recherche centré sur l'urbanité langagière.

Tout d'abord le discours enregistré des locuteurs batnéens. Ce discours a été enregistré dans différents quartiers de la ville de Batna et avec différents locuteurs (un groupe d'étudiants, des commerçants, des intellectuels, des analphabètes…) (voir annexe1). A partir des enregistrements, nous avons tenté de décrire les différentes pratiques linguistiques des différents locuteurs dans les différents quartiers de la ville pour pouvoir comprendre comment s'articulent les différents espaces urbains au sein de cette ville et aussi d'expliquer l'influence des représentations spatiales sur ces même pratiques. Sans oublier aussi qu'à travers ces mêmes enregistrements nous avons essayé de comprendre les attitudes et les perceptions des locuteurs batnéens vis-à-vis des langues, ainsi que leurs représentations spatiales. ce questionnaire a visé à dégager les représentations que les locuteurs batnéens se font de leur ville, des échanges linguistiques qui s'y déroulent. Les informations recueillies servent à compléter les enregistrements et permettent de mieux comprendre, les pratiques linguistiques de ces locuteurs et leurs stratégies adoptées dans la communication courante. La pré-enquête a consisté en plusieurs visites au sein de la communauté ciblée (la ville de Batna), visites au cours desquelles nous avons pu mener diverses activités avec les membres du groupe d'enquête : très souvent, nous nous rendons dans les différents magasins et sur les marchés de la ville de Batna pour faire soit des achats soit pour entretenir des discussions avec les gérants de ces magasins et même avec les clients. Cette démarche nous a permis tout d'abord d'approcher nos informateurs et par conséquence la délimitation de notre communauté du langage. C'est ainsi que nous avons pu identifier le groupe chaouiphone de celui des arabophones. Sans cette étape, la constitution d'un échantillon pertinent adapté à nos hypothèses aurait était impossible.

Article extrait de l'ASJP Algerian Scientific Journal Platform

CORRIGE TYPE

La méthodologie mise en œuvre dans cette recherche repose sur une approche qualitative et empirique qui combine plusieurs outils de collecte de données : **la pré-enquête, les enregistrements de discours spontanés et un questionnaire structuré.** Cette combinaison méthodologique permet d'appréhender de manière fine les pratiques linguistiques des locuteurs batnéens dans leur diversité et leur complexité, tout en les replaçant dans leur cadre spatial et social.

La pré-enquête, à travers des visites répétées dans les différents quartiers de la ville de Batna, a facilité l'identification des groupes linguistiques (arabophones, chaouiphones) et permis de cerner la communauté de langage étudiée. Cette phase exploratoire a ainsi posé les bases d'un échantillonnage pertinent, en adéquation avec les hypothèses de recherche.

Les enregistrements, réalisés dans divers lieux de la ville et auprès de locuteurs issus de milieux sociaux variés (étudiants, commerçants, intellectuels, analphabètes…), ont fourni un corpus authentique permettant de saisir les pratiques linguistiques effectives dans leur ancrage spatial. L'analyse de ces discours enregistrés met en évidence l'articulation entre les usages linguistiques et les zones urbaines spécifiques, montrant ainsi comment les représentations de l'espace influencent les choix linguistiques.

Enfin, le questionnaire a servi à recueillir les perceptions et attitudes des locuteurs à l'égard des langues utilisées dans la ville, complétant ainsi les données discursives par des éléments subjectifs essentiels à la compréhension des dynamiques sociolinguistiques locales.

En croisant ces trois outils, et les niveaux d'information, la méthodologie adoptée permet de mettre en lumière les rapports étroits entre les pratiques langagières et les contraintes de l'espace urbain, dans un contexte de plurilinguisme où coexistent l'arabe dialectal, le chaoui et le français. Ainsi, cette approche intégrée rend compte de la complexité des interactions entre langue, espace et société dans la ville de Batna.

TD 15: TELLE THEMATIQUE, TEL CORPUS : Études de Cas

CORPUS SOCIOLINGUISTIQUE

-« Cette recherche s'inscrit en sociolinguistique. C'est à travers les chansons des jeunes produites dans les villes que nous pouvons percevoir l'identité linguistique des nouvelles générations. Le Zenqawi comme nouvelle forme artistique qui commence depuis peu à prendre le devant, construit à partir du mixage de plusieurs formes linguistiques et musicales et enraciné dans un patrimoine riche, lui aussi composite par l'effet des différentes cultures et civilisations qui ont traversé le pays. Quels contextes ont été à l'origine de l'apparition de cette nouvelle identité musicale citadine ? Quels points de rencontre ce nouveau style musical opère–t-il avec les autres formes et types existant déjà dans l'espace de la ville d'Alger ? Quel rôle social (ou autre) lui assigne-t-on ? une étude sociolinguistique PAR l'analyse de chansons se réclamant du Zenqawi, dans le dessein de décrire les différentes formes d'expression du Zenqawi, dans un monde caractérisé par un plurilinguisme croissant. Le paradigme sociolinguistique est susceptible d'expliquer l'évolution du langage des jeunes et aussi de le décrire en synchronie. Le langage musical apparait alors, à l'instar du langage linguistique, comme une forme de revendication identitaire (Calvet, 1994). Notre corpus est constitué de 19 chansons classées comme relevant du Zenqawi. Etant donné que la sélection est difficile en raison de la nouveauté de ce style de chanson, nous avons procédé à une recherche par hashtag sur You Tube, en écrivant #zenqawi. En ce qui concerne la transcription des chansons, nous avons choisi le protocole ICOR qui permet de retracer les vers rapidement. La transcription phonétique de certains mots s'impose, notamment quand il s'agit des mots français algérianisés. Quant à la traduction, souvent littérale, elle permet au lecteur francophone de mieux appréhender l'analyse.

Analyse de corpus de chansons

De la protestation à la revendication identitaire

Moh Milano, Djalil Palermo, Fofa Torino, Fouzi Torino ... sont des chanteurs qui viennent des quartiers populaires d'Alger, de familles modestes, et surtout du même arrière-plan musical : des chants de stade. C'est ce qu'on relève aussi à travers les dénominations des artistes qui renvoient aux clubs sportifs italiens qui ont fait vibrer les Algériens pendant des décennies. Milano par exemple est choisi en raison de la similitude des couleurs du L'Associazione Calcio Milan, connu sous le nom Milan AC, avec l'USMA ; rouge et noir. Ces chanteurs se distinguent donc autant par leurs différences que par leurs ressemblances. Leur style de chant aspire, jusqu'à présent, à des horizons différents. Au moment où Moh Milano et Fofa Torino s'orientent exclusivement vers les chansons à caractère social, Djalil Palermo est tenté par le thème de l'amour, dans Âachkek criminel et Courage, qui se déclinent en hommage à la fidélité et la beauté de la femme, mais font aussi une critique du matérialisme caractérisant la société.

La revanche sociale est un thème central pour les chanteurs qui dénoncent le rejet d'abord des spécialistes du domaine mais aussi les accusations des responsables politiques algériens qui lient l'émigration clandestine et les chansons proposées par les jeunes chanteurs. « Je ne suis pas parti à l'étranger de manière clandestine pour demander aux jeunes de le faire. Ce que je voulais, à travers ma chanson, c'était de raconter une histoire réelle», explique Djalil Palermo. « Nous voulons raconter le vécu réel des jeunes, la vérité, c'est toute notre ambition. C'est la pauvreté, et non les chansons, qui pousse les Algériens à partir dans les barques de mort » précise Moh Milano. Dajlil Palermo chante alors Mami Nasek Delmouni

Ces chansons sont truffées de thématiques secondaires, mais qui sont aussi importantes que le premier thème et le complètent : le chômage, la pauvreté, l'enfance triste, des histoires d'échecs et de réussite aussi. Le point fort de la chanson Zenqawi est qu'elle permet, d'une part, de fredonner les histoires du peuple et, d'autre part, de proposer une source de réussite sociale pour les jeunes, ayant généralement quitté l'école trop tôt (à la fin du cycle primaire ou trois ans après), auparavant méconnus et devenus chanteurs du peuple. Ma chafouhache (ils ne l'ont pas vue) en est une illustration parfaite :

J'ai passé mon enfance dans la rue Je n'avais rien du tout Mais ce n'est pas encore fini À celui qui a été la cause On se retrouvera devant Dieu On va tous mourir un jour Je veille la nuit, j'essaye d'assurer mon avenir Je fonce, je réalise mes rêves Mon corps s'affaiblit et mon cœur dit "qu'est ce qu'il m'a pris" Un vrai petit combattant, c'est comme ça que mon père m'a élevé Ils n'ont pas vue ça, ils ne l'ont pas vu Un début difficile et on n'avait rien du tout

Ils n'ont pas vécu ça, ils ne l'ont pas vécu Un passé noir qui ne s'oublie pas (Moh Milano)

Ces chansons sont truffées de thématiques secondaires, mais qui sont aussi importantes que le premier thème et le complètent : le chômage, la pauvreté, l'enfance triste, des histoires d'échecs et de réussite aussi. Le point fort de la chanson Zenqawi est qu'elle permet, d'une part, de fredonner les histoires du peuple et, d'autre part, de proposer une source de réussite sociale pour les jeunes, ayant généralement quitté l'école trop tôt (à la fin du cycle primaire ou trois

ans après), auparavant méconnus et devenus chanteurs du peuple. Ma chafouhache (ils ne l'ont pas vue) en est une illustration parfaite :

J'ai passé mon enfance dans la rue Je n'avais rien du tout Mais ce n'est pas encore fini À celui qui a été la cause On se retrouvera devant Dieu On va tous mourir un jour Je veille la nuit, j'essaye d'assurer mon avenir Je fonce, je réalise mes rêves Mon corps s'affaiblit et mon cœur dit "qu'est ce qu'il m'a pris" Un vrai petit combattant, c'est comme ça que mon père m'a élevé Ils n'ont pas vue ça, ils ne l'ont pas vu Un début difficile et on n'avait rien du tout
Ils n'ont pas vécu ça, ils ne l'ont pas vécu Un passé noir qui ne s'oublie pas (Moh Milano)3
À la manière du Chaâbi, les chansons zenqawis, porteuses d'un message moral très fort, se présentent comme un discours de sagesse et d'adages construits par ces jeunes dans leur vie, courte mais riche en expériences, dans les quartiers défavorisés.

Article extrait de l'ASJP Algerian Scientific Journal Platform

- Comment la méthode de sélection du corpus à partir de hashtags sur YouTube influence-t-elle la représentativité et la validité des données dans une recherche sociolinguistique sur un phénomène émergent comme le Zenqawi ?
- Quelles sont les limites méthodologiques ?

1. Quel est l'objet principal de cette recherche en sociolinguistique ?
2. Pourquoi les chansons Zenqawi constituent-elles un corpus sociolinguistique pertinent ?
6. Quels éléments linguistiques peuvent traduire une revendication identitaire dans les chansons Zenqawi ?
7. Comment les artistes utilisent-ils leur langue pour exprimer un vécu collectif dans un contexte urbain ?

L'objectif principal de cette recherche est d'analyser comment le genre musical émergent Zenqawi reflète les dynamiques identitaires, linguistiques et sociales des jeunes des milieux urbains populaires en Algérie, à travers l'étude

Les chansons Zenqawi offrent un terrain riche pour la sociolinguistique car elles mobilisent des formes linguistiques hybrides (darija, français algérianisé, emprunts, expressions codées) et véhiculent des représentations sociales, identitaires et culturelles spécifiques aux jeunes citadins issus de quartiers populaires.

Les artistes du corpus proviennent majoritairement de milieux modestes des quartiers populaires d'Alger. Musicalement, ils sont issus d'un univers de chants de stade, ce qui explique l'énergie collective, le ton revendicatif et l'appartenance à une culture populaire marquée par le sport et la rue.

-Les chansons Zenqawi traitent de la pauvreté, du chômage, de la marginalisation sociale, de l'enfance difficile, des espoirs de réussite, mais aussi de l'amour et des critiques sociales. Elles expriment un vécu collectif dans une société marquée par les inégalités.

La revendication identitaire s'exprime par l'usage du parler populaire (darija), l'intégration de termes français algérianisés, les expressions locales, les alternances codiques, et une forte oralité qui déroge aux normes linguistiques officielles, valorisant ainsi un langage de la rue assumé.

-Le protocole ICOR a été utilisé pour la transcription, car il permet une représentation fidèle et structurée du discours oral, notamment utile pour repérer les vers et les séquences rythmiques dans les chansons. Ce choix facilite l'analyse fine du langage et de ses modalités expressives.

Le corpus a été constitué à partir d'une recherche ciblée sur YouTube, en utilisant le hashtag #zenqawi. Cette méthode permet de repérer les productions se revendiquant explicitement de ce genre, malgré son émergence récente et l'absence de catégorisation officielle.

On pourrait envisager une **analyse lexicale** (fréquence et variation des mots), une **analyse pragmatique** (actes de langage, positionnement énonciatif), ou une **analyse multimodale**, prenant en compte les clips vidéo, les visuels, et l'interaction avec le public sur les réseaux sociaux.

La question interroge la fiabilité du corpus (19 chansons sélectionnées par hashtag), ce qui est crucial dans toute recherche qualitative.

Elle invite à réfléchir à des méthodes d'enquêtes de terrain, pour mieux ancrer l'analyse.

TD16 : ETUDE DE CORPUS VARIES

CORPUS 01

Le chant des slogans contestataires du « hirak »

« La présente contribution se propose d'interroger les chants contestataires propres au hirak algérien, mouvement de contestation populaire amorcé en février 2019. L'intérêt sera centré, ici, sur l'une des catégories du chant que j'appellerai le « chant-slogan ». Dotés d'un aspect organisateur du mouvement, ces chants sont actualisés pour servir d'autres desseins. Usant de stratégies linguistico-discursives plurielles, le chant-slogan remplirait, dès lors, des fonctionnalités diverses organisant le mouvement et renseignant sur son ingénierie. Cet article se consacrera précisément à l'exploration de l'une de ses fonctionnalités organisationnelles. S'agissant d'un discours persuasif dans la mesure où la lutte par le langage constitue une (ré) action dont la visée globale serait d'agir pour changer un état des choses, toute une rhétorique est déployée pour atteindre pacifiquement et progressivement les objectifs escomptés.

La présente étude est à situer dans le champ de l'analyse du discours. À travers l'analyse des actes de langage (Austin, 1962) constitutifs de ces chants et la détermination de leurs visées pragmatiques, on pourrait tenter d'identifier l'auditoire visé et mieux interpréter les fonctionnalités organisationnelles du chant-slogan dans ce contexte précis de mobilisation politique. qui serait la (les) cible(s) de ces chants-slogans et surtout quelles sont les fonctionnalités organisationnelles qu'il remplit?L'étude s'appuie sur un corpus des données observables constituées à partir de la plateforme « You tube». Il est constitué d'une série de chants-slogan produits lors des manifestations hebdomadaires du hirak ayant lieu dans l'espace public algérois. Pour explorer les logiques internes de cette production langagière particulière et complexe, J'examinerai, à partir d'un corpus puisé du contexte numérique, la production et la circulation des chants-slogans dans l'espace public algérois. ».

Résumé d'article scientifique –asjp- Algerian Scientific Journal Platform

- ***Quelle technique de recueil est mise en œuvre dans cette étude***
- ***Quel serait l'apport d'une transcription orthographique et phonétique des chants dans le traitement du corpus ?***
- ***En quoi l'analyse des actes de langage (au sens d'Austin) constitue-t-elle une méthode adaptée à ce corpus .***

CORPUS 02

Supports Numériques Touristiques : Pluralité Linguistique et Traduction à L'ère du Digital

À l'ère numérique, la pluralité linguistique est cruciale pour le secteur du tourisme. Cette étude examine comment la diversité linguistique, à travers des traductions et des adaptations culturelles, influence la perception des destinations par un public multilingue. La méthodologie repose sur une analyse qualitative de contenus provenant de guides touristiques numériques, de réseaux sociaux et de forums de voyage. La méthodologie adoptée inclut une analyse qualitative de contenus provenant de guides touristiques numériques, de publications sur les réseaux sociaux, et de forums de voyage.
Les supports numériques touristiques jouent un rôle essentiel dans la manière dont les voyageurs planifient, découvrent et vivent leurs expériences. Parmi ces supports, les applications de voyage se distinguent par leur capacité à améliorer l'expérience à chaque étape du parcours. Elles offrent des fonctionnalités variées, allant de la réservation de vols et d'hébergements (Booking.com) à la navigation (Google Maps, 2023) et la découverte d'activités locales (Culture Trip). Les sites web des destinations, gérés par des offices de tourisme, fournissent des informations complètes sur les attractions et événements (Visit Finland, 2023). Les réseaux sociaux et le contenu généré par les utilisateurs deviennent des sources d'inspiration incontournables, permettant aux voyageurs de partager leurs expériences et d'influencer les décisions des autres (WebFX, 2024). Enfin, les vidéos et le contenu audiovisuel offrent une immersion visuelle dans les destinations, suscitant l'envie de voyager et fournissant un aperçu authentique des lieux à visiter (Buzz, 2024).

Une variété de supports numériques touristiques, tels que des applications et des blogs de voyage multilingues, des sites web de destinations de voyage, des guides touristiques numériques, des forums de voyage et des contenus sur les réseaux sociaux. Dans l'examen de diverses applications de voyage, plusieurs exemples mettent en lumière la manière dont la pluralité linguistique est valorisée. De plus, cette pluralité linguistique influence significativement la qualité et l'efficacité des traductions. Voici quelques études de cas à cet égard : • Cas 1 : Booking.com L'application de réservation Booking.com illustre la pluralité linguistique en adoptant une méthode de traduction automatisée, complétée par une révision effectuée par des experts humains. Par exemple, dans la description : "Nos chambres offrent une cozy ambiance pour 136 Meriam BENLAKDAR et Ahlem HAL un séjour relaxant.", le mot "cozy" est souvent utilisé en français pour décrire une ambiance chaleureuse, ajoutant une touche moderne à la description. En outre, dans la phrase "Profitez de notre room service disponible 24h/24.", l'usage de "room service" est courant en français, apportant une connotation internationale et une familiarité pour les voyageurs. Le terme "workshop", dans la phrase "Participez à nos workshops de cuisine locale.", est intégré dans le vocabulaire français, soulignant une tendance à utiliser des termes anglais pour désigner des activités interactives. Le terme "Lounge bar" dans la phrase "Dégustez nos plats au lounge bar élégant.", est un terme qui évoque une ambiance moderne et décontractée, souvent plus percutant pour attirer une clientèle jeune. Ainsi, la description des hôtels illustre la richesse de la pluralité linguistique et les nuances culturelles, en intégrant des éléments spécifiques qui parlent à un public diversifié. Enfin, des références telles que "spécialités méditerranéennes et andalouses" et "touches arabes et andalouses" évoquent une fusion culturelle, soulignant l'histoire et l'architecture de la région. Par ailleurs, en comparant les traductions des descriptions d'hôtels du français vers l'anglais, on observe que certaines nuances culturelles peuvent être perdues. Par exemple, des termes comme "chambres cocooning" sont traduits par "cozy rooms", ce qui, bien que correct, ne transmet pas entièrement l'aspect intime et chaleureux évoqué dans la langue source.
Cette analyse souligne l'importance de conserver les connotations culturelles spécifiques lors de la traduction. En outre, l'utilisation de termes comme "buffet" et "à la carte" reflète des concepts culinaires qui peuvent varier d'une culture à l'autre. L'expression française "Chambres avec vue sur la mer" peut évoquer une expérience romantique et sereine, tandis qu'en anglais, "ocean view" est plus factuel, sans la même connotation de tranquillité. De même, l'expression "Vue imprenable" en

français souligne non seulement la beauté, mais aussi l'importance de la perspective dans l'expérience. "Breath taking view" en anglais, bien que poétique, peut sembler plus descriptif et moins ancré dans un contexte émotionnel ou culturel. L'adjectif "Convivial" dans l'expression "Ambiance conviviale" implique une chaleur humaine et une interaction sociale, alors que "friendly atmosphere" peut se limiter à une simple courtoisie sans la profondeur relationnelle. L'expression "Espace de bien-être" implique une approche holistique de la santé, intégrant le corps et l'esprit, alors que "wellness area" en anglais peut se concentrer davantage sur les aspects physiques, comme les installations de fitness. De plus, dans l'expression "Spa de luxe avec soins personnalisés", le terme "soins personnalisés" en français évoque une attention méticuleuse aux besoins individuels du client, tandis que "Luxury spa personalized treatments" en anglais peut sembler plus standardisé. La nuance culturelle ici réside dans l'idée que le français valorise une expérience plus intime et sur mesure.
Sur Instagram et Facebook, les utilisateurs intègrent souvent des hashtags dans plusieurs langues pour atteindre un public plus large. A titre d'exemple, nous citons cette publication "Exploring the beauty of the Sahara! #Sahara #Désert #Aventure". Ainsi, la pluralité linguistique se manifeste clairement à travers l'utilisation de hashtags en anglais et en français illustrant une volonté d'inclure des locuteurs de différentes langues, reflétant ainsi une diversité culturelle. Cela permet aux utilisateurs de découvrir le contenu, A cet effet, ils peuvent se servir de la traduction automatique. ❖ Légendes Multilingues Les publications incluent souvent des légendes dans plusieurs langues pour toucher divers groupes culturels. Par exemple : la légende : "La beauté de la Casbah d' Alger est incroyable! / The beauty of the Casbah of Algiers is amazing!", cela montre une ouverture à la diversité et une reconnaissance des différentes cultures qui composent la communauté des voyageurs qui commentent souvent dans des langues différentes, créant un dialogue multilingue. Nous citons ce commentaire en Français : "C'est magnifique! Je rêve d'y aller un jour." , et la réponse en Anglais : "You should definitely visit! Ce type d'échange illustre une dynamique culturelle et favorise la compréhension interculturelle, permettant aux utilisateurs de partager des expériences de voyage.

Source : ASJP

CORRIGE TYPE

CORPUS 01

TECHNIQUES DE RECUEIL DES DONNEES

Le corpus repose sur des **enregistrements** de manifestations Hirak (principalement des clips postés sur YouTube). La collecte se fait par téléchargement des vidéos des manifestations hebdomadaires, en ciblant en priorité la période du vendredi à Alger. Chaque vidéo est annotée par ses métadonnées principales (date, lieu exact de la marche, nombre estimé de manifestants, langues dominantes chantées, présence de banderoles ou slogans visibles).

Il s'agit de sélectionner un échantillon diversifié : par exemple plusieurs vendredis consécutifs, différents quartiers (centre, Bab El-Oued, etc.) et si possible différents meneurs de chants. Cette stratégie suit l'exemple de Boudjemaa et Oulebsir (2023) qui ont analysé un corpus cadré sur les six premières manifestations. L'échantillonnage proposé devra couvrir au moins plusieurs

dizaines de chants-slogans (par exemple 10 à 20 chants de manifestants filmés), afin d'assurer une certaine représentativité sans prétendre à l'exhaustivité.

Transcription des données de corpus

Chaque chant-slogan étant oral, une transcription précise est essentielle. Nous proposons une transcription **orthographique en graphie latine**, adaptée au dialecte algérien. On écrira par exemple *ramenez* ou *tarhal* en toutes lettres pour refléter les formes prononcées. En complément, une transcription **phonétique en API** permettra de noter les sons qui n'existent pas en alphabet latin. Par exemple, dans le corpus de [15] les auteurs ont transcrit six chants en graphie latine pour la lisibilité et ont utilisé l'API pour /ح/ ,/خ/ ,/ع/ (symboles [ʾ], [x], [ħ] respectivement). On adoptera de même des conventions explicites (p. ex. [ʕ] ou [ʾ] pour /ع/, [x] pour /خ/, [ħ] pour /ح/).

L'intérêt de cette double transcription est double : d'une part elle fixe rigoureusement la forme lexicale et phonétique des slogans, en distinguant l'« accent action » (intonation, allongements) du contenu verbal; d'autre part, elle facilite l'analyse linguistique (traitement automatique du corpus, recherche de motifs, comparaison de variantes dialectales).

Modalité d'Analyse des Actes De Langage (Austin)

Nous adoptons une approche pragmatique : chaque phrase chantée est traitée comme un **acte de langage** au sens d'Austin (1962) et Searle. Cette méthode est adaptée car les chants-slogans cherchent avant tout à **agir** : informer, exhorter, menacer ou mobiliser l'auditoire. L'analyse identifie pour chaque énoncé son **illocution** (ordre, demande, déclaration, expression de sentiments, etc.) et son **effet perlocutoire** attendu (motiver, conscientiser, rassurer, etc.). Comme le notent Ouaras et al., cette démarche permet « d'identifier l'auditoire visé et mieux interpréter les fonctionnalités organisationnelles du chant-slogan ».

Par exemple, le slogan *« Ramenez la BRI, ramenez la BRG… il n'y a pas de 5e, oh Bouteflika »* se compose de deux segments : l'impératif « Ramenez la BRI… » est un acte directif/demande adressé aux forces de l'ordre, et son perlocutoire est clairement de défier l'autorité (« vous n'avez pas le contrôle, on continue »). Comme l'explique Oulebsir, l'impératif exprime moins un ordre qu'une défiance qui construit un *ethos* collectif de bravoure chez les manifestants. Le second segment « il n'y a pas de 5e, oh Bouteflika » est un acte

déclaratif (évidence/volonté), qui, perlocutoirement, affirme la détermination du groupe à contester le 5e mandat.

Nous concevons un protocole d'annotation en deux volets :

(i) pour l'acte illocutoire, on utilisera une classification type Searle (représentatives, directives, commissives, expressives, déclaratives) ;

(ii) (ii) pour l'effet perlocutoire, on codera l'intention générale (mobilisation du public, légitimation du mouvement, pacification, dénonciation, etc.). Cette annotation sera effectuée manuellement par au moins deux analystes pour assurer la fiabilité. Les extraits annotés illustratifs sont présentés dans le tableau suivant.

Tableau : Modalité d'analyse des actes de langage en corpus01

Extraits	**Acte illocutoire (catégorie)**	**Effet perlocutoire (but recherché)**
« Ramenez la BRI, ramenez la BRG… »	Directif implicite (impératif) – *demande/défi*	Défier l'autorité et affirmer la détermination collective
Il n'a pas de 5ème mandat Bouteflika.	-Déclaratif – affirmation/certitude	Renforcer la détermination. Rejeter le 5eme mandat
chaque vers chanté est accompagné de son acte de langage principal et de la fonction pragmatique associée. Cette annotation met en évidence comment les slogans agissent sur l'auditoire (par exemple, en créant un lieu commun de défiance ou en rappelant un objectif revendicatif).		

CORPUS 02

Supports Numériques Touristiques : Pluralité Linguistique et Traduction à L'ère du Digital

Synthèse

Cette étude met en lumière l'importance cruciale de la pluralité linguistique et de la traduction dans le secteur du tourisme numérique. À l'ère où les interactions se multiplient sur des plateformes digitales, il est impératif que les acteurs du tourisme reconnaissent et intègrent la diversité linguistique dans leurs stratégies de communication. Les analyses réalisées démontrent que des traductions précises et culturellement adaptées favorisent une meilleure

compréhension et une expérience enrichissante pour les voyageurs multilingues. Les défis identifiés, tels que la perte de nuances culturelles dans les traductions et les barrières linguistiques, soulignent la nécessité d'une approche proactive. En adoptant des bonnes pratiques de traduction et en sensibilisant les professionnels du tourisme à l'importance de la diversité linguistique, il est possible d'améliorer l'engagement des utilisateurs et de promouvoir une expérience touristique authentique. Enfin, à l'avenir, il sera essentiel de continuer à développer des outils et des ressources qui soutiennent la pluralité linguistique, tout en respectant et en valorisant les cultures locales. En intégrant ces éléments dans leurs offres, les acteurs du secteur pourront non seulement enrichir l'expérience des voyageurs, mais également contribuer à un tourisme plus inclusif et durable.

TD17 : DU MEMOIRE ACADEMIQUE AU MEMOIRE STRAT UP : DES CORPUS A ADAPTER

Dans le cadre des nouveaux dispositifs universitaires (arrêtés 1275/008). L'enjeu consiste à opérationnaliser les objets linguistiques en produits ou services innovants (applications, plateformes, outils d'analyse, solutions éducatives, etc.).

Proposez une transformation structurée des thématiques ci-dessous (sujets de recherche des mémoires académique pour l'année universitaire2025-2026) en idées de mémoire entreprise – économique ou start-up, avec une logique claire explicite qui répond aux éxigences de la spécialité-linguistique-:

L'idée ou le Phénomène étudié → Valeur ajoutée et objectifs → potentiel économique – entrepreunarial.

Intitulé du mémoire	**suggestions pour un mémoire Start-up**
« Analyse discursive de la polyphonie identitaire dans « Ciel de l'Agha » de Hacène Tebbal	Start-up proposée : Plateforme d'analyse narrative assistée par IA - Idée : outil numérique qui analyse automatiquement les discours littéraires (identité, exil, mémoire). - Utilisateurs : étudiants, chercheurs, enseignants. - Valeur : gain de temps + visualisation des structures discursives.
Étude diachronique de l'anthroponymie à Bordj Bou Arreridj (1980–2020) : approche sociolinguistique	Start-up : Application d'analyse des prénoms et identités culturelles - Idée : plateforme qui explique l'origine, l'évolution et la dimension identitaire des prénoms. - Marché : grand public + généalogie + sociologie.
« l'acquisition du langage chez l'enfant en contexte plurilingue algérien	- Start-up : Application éducative plurilingue pour enfants algériens - Idée : app interactive pour développer langage (arabe/français/anglais). - Public Cible : parents, écoles privées. -
Représentations du passage du français à l'anglais dans la formation médicale en Algérie.	Start-up : Plateforme de formation linguistique spécialisée (anglais médical) - Idée : modules interactifs pour étudiants en médecine. - Glossaire, simulation de cas, prononciation.
Étude lexico-sémantique des dénominations des produits de grande consommation en Algérie.	➞ Start-up : Agence de Dénominations commerciales (produits) basée sur l'analyse linguistique • création de noms de marques adaptés au marché algérien. • approche scientifique (lexico-sémantique). • Clients : entreprises, start-up.
Étude linguistique des dénominations des enseignes de restaurations à Bordj Bou Arréridj : approche onomastique et sociolinguistique	- Concept pour start-up : Branding linguistique pour les commerces Concept : accompagnement des professionnels dans la création et l'optimisation des noms d'enseignes afin d'en renforcer l'attractivité, la mémorisation et l'ancrage culturel.
Etude linguistique du discours journalistique : cas des néologismes dans les chroniques de Hakim lalaam.	Concept à base de néologisme Base de données dynamique des nouveaux mots.

CONCLUSION GENERALE

Au terme de ce cours consacré à la linguistique de corpus, il apparaît clairement que cette approche constitue aujourd'hui un pilier fondamental de la recherche en sciences du langage. En plaçant les données empiriques au cœur de l'analyse, elle permet de dépasser les limites des approches intuitives et d'ancrer l'étude des phénomènes linguistiques dans des usages réels et contextualisés. Cette orientation méthodologique répond aux exigences contemporaines de scientificité, fondées sur la vérifiabilité, la reproductibilité et la rigueur des analyses.

L'un des apports majeurs de la linguistique de corpus réside dans sa capacité à articuler différentes dimensions de l'analyse linguistique. En effet, les corpus offrent un terrain d'observation privilégié pour étudier des phénomènes relevant aussi bien de la syntaxe, du lexique, de la sémantique que de la pragmatique ou de la sociolinguistique. Ils permettent également d'explorer les pratiques discursives dans leur diversité, en tenant compte des contextes sociaux, culturels et institutionnels dans lesquels elles s'inscrivent. Ainsi, la linguistique de corpus favorise une approche intégrative, capable de rendre compte de la complexité des faits de langue.

Par ailleurs, ce module a mis en évidence l'importance des différentes étapes de la recherche empirique. De la formulation de la problématique à l'interprétation des résultats, en passant par la constitution et le traitement du corpus, chaque phase nécessite une réflexion méthodologique approfondie. Le chercheur doit faire des choix éclairés concernant la sélection des données, les outils d'analyse ou encore les cadres théoriques mobilisés. Ces choix conditionnent la validité des résultats obtenus et la pertinence des conclusions formulées.

Les travaux dirigés ont joué, dans ce cadre, un rôle essentiel en permettant aux étudiants de mettre en pratique les concepts abordés en cours. À travers la réalisation d'exercices concrets, ils ont pu expérimenter les différentes étapes de la recherche sur corpus, depuis la collecte des données jusqu'à leur analyse. Cette dimension pratique est indispensable pour développer une véritable compétence méthodologique, fondée sur l'expérience et la réflexion critique.

En outre, l'accent mis sur la diversité des approches analytiques a permis de montrer que la linguistique de corpus ne se réduit pas à une seule méthode, mais qu'elle constitue un champ ouvert, traversé par des perspectives théoriques variées. L'approche énonciative, par exemple, met l'accent sur les marques de subjectivité dans le discours, tandis que l'approche lexicométrique privilégie l'analyse statistique des occurrences. De même, les approches dialogiques et sémiologiques permettent d'explorer les interactions entre les discours et les

systèmes de signes. Cette pluralité constitue une richesse, dans la mesure où elle offre au chercheur une palette d'outils pour appréhender les données sous différents angles.

Enfin, la question de l'adéquation entre la thématique de recherche et le corpus a été au cœur de ce module. Comme le souligne l'axe « Telle thématique, tel corpus », le choix des données ne peut être dissocié des objectifs de recherche. Un corpus pertinent est celui qui permet de répondre de manière précise et argumentée à une question donnée. Cette exigence implique une réflexion préalable sur les critères de sélection, la représentativité et la pertinence des données.

Dans une perspective plus large, la linguistique de corpus s'inscrit pleinement dans les évolutions actuelles des sciences humaines et sociales, marquées par le développement des technologies numériques et l'accès massif aux données. Les outils informatiques offrent aujourd'hui des opportunités inédites pour la collecte, le traitement et l'analyse des corpus, ouvrant ainsi de nouvelles perspectives de recherche. Toutefois, ces avancées techniques ne dispensent pas le chercheur d'une réflexion critique sur ses méthodes et ses résultats.

En définitive, ce module a permis de mettre en lumière les enjeux, les méthodes et les perspectives de la linguistique de corpus. Il a contribué à former des étudiants capables de mener des recherches rigoureuses, fondées sur des données empiriques, et de mobiliser des outils analytiques adaptés à leurs problématiques. Plus qu'un simple ensemble de techniques, la linguistique de corpus apparaît ainsi comme une véritable démarche scientifique, au service d'une meilleure compréhension des langues et des discours.

Références bibliographiques

Benveniste, É. (1966). *Problèmes de linguistique générale 1*. Paris : Gallimard.

Benveniste, É. (1974). *Problèmes de linguistique générale 2*. Paris : Gallimard.

Blanchet, P. (2000). *La linguistique de terrain : Méthodes et théories*. Rennes : Presses universitaires de Rennes.

Blanche-Benveniste, C. (1996). De l'utilité du corpus linguistique. *Revue française de linguistique appliquée*, 1(2).

Charaudeau, P., & Maingueneau, D. (2002). *Dictionnaire d'analyse du discours*. Paris : Seuil.

Dalbera, J.-P. (2002). Le corpus entre données, analyse et théorie. *Corpus*, (1). http://corpus.revues.org/document10.html

Dubois, J., Guespin, L., Giacomo, M. C., Marcellesi, J.-B., & Mével, J.-P. (1994). *Dictionnaire de linguistique et des sciences du langage*. Paris : Larousse.

Ducrot, O. (1984). *Le dire et le dit*. Paris : Minuit.

Ducrot, O., & Schaeffer, J.-M. (1995). *Nouveau dictionnaire encyclopédique des sciences du langage*. Paris : Seuil.

Guilhaumou, J. (2002). Le corpus en analyse du discours : perspective historique. *Corpus*, (1).

Habert, B., Nazarenko, A., & Salem, A. (Éds.). (1997). *Les linguistiques de corpus*. Paris : Armand Colin.

Jacobson, M. (2005). *Corpus oraux en linguistique de terrain*. https://hal.science/hal-00004909

Kennedy, G. (1998). *An introduction to corpus linguistics*. London : Longman.

Kerbrat-Orecchioni, C. (1998a). *La conversation*. Paris : Armand Colin.

Kerbrat-Orecchioni, C. (1998b). *Les interactions verbales : Approche interactionnelle et structure des conversations* (Tome 2). Paris : Armand Colin.

Khelifi Hadria, H. (2016). La lexicométrie : un outil efficient pour l'analyse du discours. In *Actes du colloque international Médias numériques, langues, discours, pratiques et interculturalité* (pp. 151–161). Agadir.

Maingueneau, D. (2005). L'analyse du discours et ses frontières. *Marges linguistiques*, (9), 64–75.

Mayaffre, D. (2002). Les corpus réflexifs : entre architextualité et hypertextualité. *Corpus et recherches linguistiques*, (1).

Mayaffre, D. (2005). Les corpus politiques : objet, méthode et contenu. Introduction. *Corpus et recherches linguistiques*, (4).

Moirand, S. (2007). *Les discours de la presse quotidienne : Observer, analyser, comprendre*. Paris : PUF.

Mondada, L. (2005). L'analyse de corpus en linguistique interactionnelle : De l'étude de cas singulier à l'étude de collections. In A. Condamines (Éd.), *Sémantique et corpus* (pp. 77–108). Paris : Hermès.

Rastier, F. (2005). Enjeux épistémologiques de la linguistique de corpus. In G. Williams (Éd.), *La linguistique de corpus*. Rennes : Presses universitaires de Rennes.

Salem, A. (1993). *Méthodes de la statistique textuelle* (Thèse de doctorat). Université Paris III.

Traverso, V. (1999). *L'analyse des conversations*. Paris : Nathan.