

**Utilisation du corpus comme une ressource
et des outils de recherche**

Présenté par : Ghalia moussa El shikhi

Use of study samples as a source and means for search

We present in this research definitions of the basic parts as well as the tools that help in its study, such as the locus and the electronic sources.

We study the types of the basic parts, explaining the corpus of the types such as the basic parts written or text.

And we provide examples of the basic parts of the field the language of the name "Outreau" and the multiple meaning and locations in the sentences.

In this research we examine the basic parts of the literature through the story of the mute to Emile Zola and the indirect speech.

We find through our study that the basic parts of corpus.

استخدام عينات الدارسة كمصدر وأداة للدارسة

Corpus هو الأجزاء الأساسية المراد دراستها في البحث العلمي .

في البداية نقدم تعريفات لمصطلح الأجزاء الأساسية في البحث وأيضاً الأدوات التي تساعد في دارسته كالفاموس والمصادر الالكترونية لشرح الكلمات والمعاني .

ندرس في هذا البحث الأجزاء الأساسية موضحاً ما يحتويه كل نوع .

ونقدم أمثله عن الأجزاء الأساسية في مجال علم اللغة كاسم العلم اوترو - وتعدد معانيه ومواقعه في الجمل .

نتطرق أيضاً إلى الأجزاء الأساسية المراد دراستها في البحث اللغوي الذي يستخدم الأدب من خلال إحدى روايات الكاتب الفرنسي إميل زولا والكلام غير المباشر وكيفيه دراسته.

نجد خلال دارستنا للأجزاء الأساسية المراد دراستها أن كوربوس أو العينة تحتوى على مضمون يكون إما له علاقة بعلم الأصوات أو تحليل المدلول اللغوي ، أو تحليل اجتماعي اقتصادي ، أو لغوي أو نصي وغيرها من العلوم .

**(Utilisation le corpus comme une ressource
et des outils de recherche)**

1 – Introduction :

La conception de l'objet corpus est comme une question cardinale pour les pratiques linguistiques contemporaines, le corpus constitue d'images, de textes, ou de phrases pour l'analyse ou l'étude, et l'auteur propose des outils innovants et de source pour l'analyse graphique et l'analyse textuelle, il permet aux linguistes et aux chercheurs de comprendre les évolutions de l'écriture dans l'environnement multilingue. et dans cette étude nous allons présenter le corpus et expliquer son importance.

A travers cette étude, nous étudions le corpus dans la recherche scientifique en présentant des exemples du corpus dans la linguistique « le nom Outreau » et du corpus dans la grammaire stylistique (le discours rapporté indirect dans le roman de Zola « L'Assommoir »)

2-Définitions :

D'abord nous citons certaines définitions du corpus :

1-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)¹. « Sinclair cité par Habert et a, 1997 :144 »

2-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)² «Rastier François, 2004 « Doxa et lexique en corpus-pour une sémantique des idéologies.

3-Le corpus peut se définir comme un recueil de données :

-Outils et (corpus outillés) (corpus écrits) :

-Le dictionnaire en ligne TLFI :

-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.

3–Types de corpus :

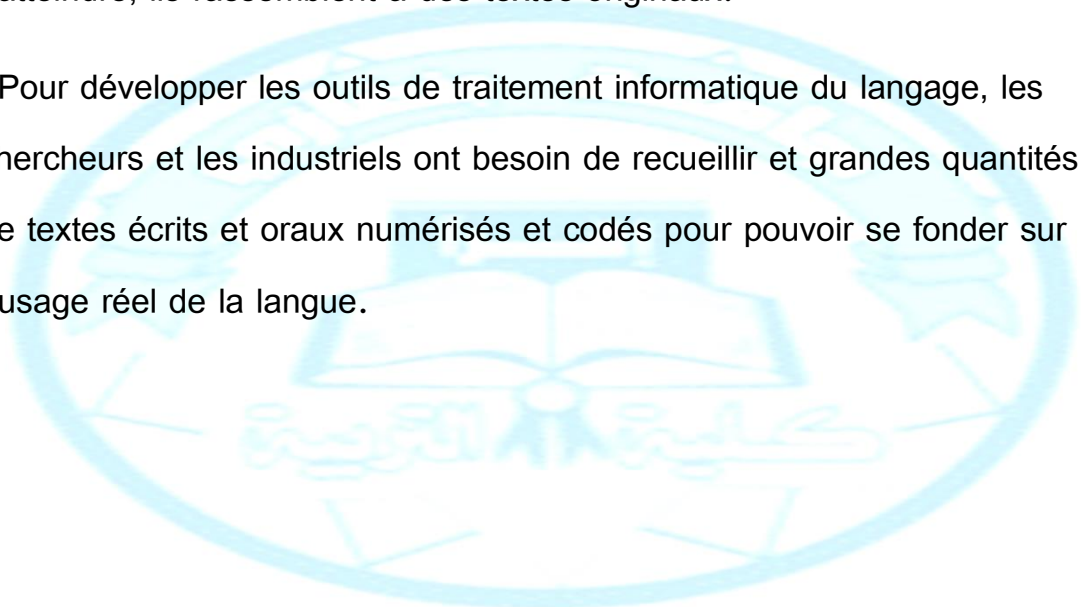
–Il existe plusieurs types de corpus ;

1–Le corpus (sacs de mots) : généralement utilisé par les lexicographes.

2–Les corpus des grammairiens : généralement font de phrases hors contexte mais attestant de la bonne formation des énoncés.

3–Les corpus (archives textuelle) : visant l'exhaustivité sans jamais l'atteindre, ils rassemblent à des textes originaux.

–Pour développer les outils de traitement informatique du langage, les chercheurs et les industriels ont besoin de recueillir et grandes quantités de textes écrits et oraux numérisés et codés pour pouvoir se fonder sur l'usage réel de la langue.



Les ressources linguistiques sont des recueils de données informatisées de production langagières écrits ou parlées.

Les corpus écrits sont les plus fréquents.

3-1-Types des corpus écrits :

Les corpus écrits ont les types suivants :

- Corpus bruts : textes écrits ou transcriptions écrites de productions orales comme on en obtient grâce au système de codage TEI (Texte encoding initiative).
- Corpus annotés ou enrichis : avec des indications relatives à la structure du texte, aux catégories morphosyntaxiques ou sémantiques.
- Corpus alignés : avec un contenu traduit en plusieurs langues.

Pour le concordancier bilingue, il permet à l'utilisateur d'extraire toutes les occurrences d'une chaîne de caractères déterminée (mot isolé, groupe de mots, expressions, etc.) à partir de la base de données d'une mémoire de traduction (MT) et de visualiser ces mêmes occurrences dans leur contexte immédiat.

La fonction de concordance fait de la technologie des MT un instrument précieux d'aide à la traduction pour les textes qui, contrairement aux révisions et aux mises à jour, ne contiennent qu'un nombre très limité de segments inférieure à celle de la phrase (propositions, expressions, etc.).

–Outil informatique (logiciel) qui permet d’extraire d’un texte des occurrences accompagnées de leur contexte : concordances.

–Un outil n’est pas intelligent, c’est l’utilisateur qui l’est !

–Non pertinence de la notion de « mot », de « phrase ».

–Une phrase est un ensemble de mots concourant à exprimer un ensemble d’idées, un sens suivi et complet. Considérée grammaticalement, la phrase se compose du sujet et de ses dépendances, du verbe et de ses compléments directes, indirects, circonstanciels, et peut renfermer des mots appartenant à toutes les parties du discours ; ces mots sont soumis à certaines règles de construction et de syntaxe.

Pour les mots, on appelle mot un son ou un ensemble de sons qui exprime une idée distincte C.à.d. qui exprime une représentation, une sensation ou une conception.

–Scientext : une base de données de textes scientifiques contemporaines ; outillée.

–Frantext : une base de données (et non un corpus) ; outillée.

La toponymie est la science qui étudie les noms de lieux ou toponyme. Elle propose de rechercher leur signification, leur étymologie, leur évolution et leur impact sur les sociétés.

5-Comment étudie-t-on le corpus ?

- Présenter les textes ou les phrases.
- Lire attentivement tous les phrases du corpus avant de commencer étudier.
- Assurer d'avoir bien compris le sens des phrases, De quoi s'agissent-ils ? Qui parle ? À qui ? Quel est le type de phrase ?etc.
- Souligner les mots importants.
- Relire les phrases pour repérer les passages des phrases qui nous intéressent pour répondre à la question de départ ou la problématique.
- Ensuite on commence à analyser les corpus, par exemple le nom « Outreau »

6-Travail sur le nom propre « Outreau » :

Le cas des toponymes ;

1-Point de départ : certains noms propres (NP) ont plusieurs sens, et certains sens de noms propres (NP) changent.

2-Point de départ (corrolaire) : le NP a du sens.

Des exemples :

1-La Martinique *s'indigne, crie, se soulève*...les Martiniquais veulent vivre mieux. Mais surtout, bien. (..)Aujourd'hui, la Martinique et la

Guadeloup, *aux voix* trop souvent ignorées, veulent être entendues.

(Mireille Fanon-Mendès France, Polities, 5 mars 2009)1.

2–Un millier des exposés internationaux sont réunis cette semaine à Vienne pour faire le point sur les *conséquences sanitaires et socio-économiques* de Tchernobyl. (Archives le monde, 1995–1996)2.

3–Marseille s'incline à Auxerre (Yahoo23/12/09)3.

4–Le verdict d'Outreau ne confirme-t-il pas l'impression d'un fiasco judiciaire ? (le Figaro 3/7/04)1.

–Projet de recherche : observer en corpus le changement de sens d'un NP.

–Le NP Outreau, nom d'une ville du nord de la France (Pas-de-Calais).

–Outreau, lieu d'événements dramatiques, rapportés par les médias ; littérature abondante (presse, ouvrage de Florence judiciaire)2.

Ex : pour éviter un nouvel Outreau (les Echos, 2/12/05, en titre)3.

7–Constitution du corpus :

Corpus de presse, partition du corpus.

En 2001–2002 : on parle de réseau de pédophilie.

Ex : Outreau : un réseau d'ouvriers et de notables viols, meurtres, pédophilie : les fantômes d'Outreau (le Figaro, 14/01/02)4.

8–Les analyses : comment rentrer dans ce corpus ?

Utilisation de concordance Antconc : environnement, fréquences et rang de fréquence, comparaisons entre sous-corpus.

-le contexte :

Syntaxique et distributionnel : palier du syntagme ou de la phrase.

-après Outreau.

-habiter Outreau.

-l'environnement du mot réseau ; environnement des noms « d'actants » de l'affaire d'Outreau.

-le réseau d'Outreau vs un réseau à Outreau.

La constitution du corpus, c'est-à-dire sur les possibilités et les conséquences de la clôture du champ d'investigation.

Il est désormais acquis qu'un corpus n'est pas un ensemble de données langagières en vrac mais des données qu'on décide de regrouper pour une étude particulière.

Le travail sur le corpus, qui permet notamment sa viabilisation informatisée, consiste donc essentiellement à former un sens à partir des manipulations qui font alterner vues analytiques et synthétiques sur les significations, multipliant ainsi les aspects de ces dernières tout en leur assignant des définies.

9–Le corpus et la littérature française :

Le corpus en littérature réfléchit à l'enseignement de la littérature à partir de la question du corpus.

Le corpus consiste à sélectionner des œuvres (lues ou à lire) ; ce qui revient à découper, dans le champ infini des œuvres que l'on pourrait donner à lire, un espace qui est aussi un miroir réfractant les valeurs de la société ou de l'individu, le corpus est établi, suivant les contextes, à partir des progressions des enseignants ou formateurs d'enseignants.

Nous allons présenter un exemple du corpus dans la littérature française tiré du roman « Assommoir » d'Emile Zola, ici le corpus compose des phrases qui étudient le discours rapporté au passé indirect de côté grammatical (la grammaire stylistique) et sémantique :

1–la jeune femme expliqua qu'elle attendait une personne. P94

Verbe introducteur : expliquer.

Propos rapportés : qu'elle attendait une personne.

Type de phrase : déclarative.

2-il lui demanda encore si elle voulait. P95

Verbe introducteur : demander.

Propos rapportés : elle voulait.

Type de phrase : interrogative.

3-elle lui jurait qu'elle ne souffrait plus du tout. P154

Verbe introducteur : jurer.

Propos rapportés : elle ne souffrait plus du tout.

Type de phrase : déclarative.

4-il racontait qu'elle lui faisait des chatouilles. P200

Verbe introducteur : raconter.

Propos rapporté : elle lui faisait des chatouilles.

Type de phrase : déclarative.

5-on lui criait que c'était son tour. P277

Verbe introducteur : crier.

Propos rapportés : c'était son tour.

Type de phrase : déclarative.

6-il raconta comment ils s'étaient rencontrés rue-Rochechouart. P287

Verbe introducteur : raconter.

Propos rapportés : ils s'étaient rencontrés rue-Rochechouart.

Type de phrase : déclarative.

7-on lui disait que c'était de bon cœur. P294

Verbe introducteur : dire.

Propos rapportés : c'était de bon cœur.

Type de phrase : déclarative.

8-Boche disait que Nana et Pauline étaient des femmes. P380

Verbe introducteur : dire.

Propos rapportés : Nana et pauline étaient des femmes.

Type de phrase : déclarative.

9-il lui demanda furieusement d'où venaient ces rubans. P425

Verbe introducteur : demander.

Propos rapportés : où venaient ces rubans.

Type de phrase : interrogative.

10-elle déclara qu'elle voulait travailler chez elle. p445

Verbe introducteur : déclarer.

Propos rapportés : elle voulait travailler chez elle.

Type de phrase : déclarative.

(Emile Zola, L'Assommoir, paris, Charpentier, 1877.)¹

Le discours rapporté dans la littérature (ou la grammaire stylistique) est un objet d'étude, mais les méthodes seront en partie transposables à d'autres genres et à d'autres disciplines des sciences humaines (histoire, sociologie,...etc.)

Le corpus permet de poser la question de la représentation du littéraire et plus profondément d'une évolution majeure.

Donc les corpus sont des outils indispensables pour la recherche en tant que terrains d'expérimentation, pour tester exhaustivement des hypothèses, qu'en tant que terrain de découverte pour en susciter des nouvelles.

10-Conclusion

Nous trouvons que le corpus constitue un objet fondamentalement nouveau par rapport à celui que l'on dévolut souvent à la linguistique (la langue), principalement car il est limité.

On trouve encore que le corpus est caractérisé par un certain nombre qui permet la discussion, il se définit par ordre déterminé.

Le contenu des corpus est déterminé par l'objectif de la recherche : phonétique, phonologie, analyse des interactions, analyse de la syntaxe et de la morphologie, analyse des variations en fonction de classements

socioéconomiques ou autres ; par les genres et situation de parole, parole spontanée ou sollicitée, monologues, dialogue, conversations, parole privée ou publique ,formelle ou familière, avec des explications, argumentations ou narrations ;et par le type de communication face à face ; radio, télévision ou téléphone.

Donc l'utilisation du corpus se complète de l'emploi de bases de données ou de dictionnaires, il donne des pistes sur la manière de traduire une référence culturelle et des ressources textuelles dans différentes langues.

Bibliographie

- **archive le monde, 1995–1996.**
- **Emile Zola, L'Assommoir, paris, charpentier ,1877.**
- **François Rastier, Doxa et lexique en corpus–pour une sémantique des « idéologies ».**
- **Les Echos, 2/12/2005, en titre.**
- **Le Figaro, 14/01/2002.**
- **Le Figaro 3/7/2004.**
- **Mireille Fanon–Mendès France, Politis, 5 mars 2009.**
- **Sinclair cité par Habert et a, 1997 :144.**