

## Point de vue sur l'analyse des corpus oraux

**Haif Nour El Houda**

**Pr.Dakhia Abdoulouahab**

**Laboratoire FEU**

**Université Kasdi Merbah -Ouargla-**

### **Résumé :**

Les études et les recherches scientifiques varient en fonction de la nature du corpus à analyser, ce qui détermine aussi la méthode et les grands axes d'expérimentation effectuée. La contribution que nous proposons interroge un certain nombre d'éléments majeurs concernant le corpus oral, elle situe précisément : d'une part l'étape de la collecte des données dans laquelle s'inscrit l'opération de transcription, d'autre part ce qui fait la particularité du traitement de ce genre de données orales.

Il convient donc que le processus d'analyse ne dépend pas uniquement de l'enregistrement de l'événement communicatif produit par les locuteurs, si bien qu'il peut paraître fort compliqué, car nous nous retrouvons confrontés aux problèmes de constitution de corpus.

En fait, c'est une collecte de différents discours dits oraux que nous visons à étudier. D'emblée nous mettons l'accent sur ce genre, sans faire recours à tous les sous-types car il n'est pas abordable d'entamer toutes les spécificités de ce genre lors d'une seule intervention.

Quelles que soient les esquisses d'une telle étude interdisciplinaire s'inscrivant dans le contexte des genres oraux, le chercheur est obligé de mener un processus de sélection soigneusement en fonction des exigences de sa

recherche, ainsi de soumission et de classement, pour obtenir des données cohérentes.

Les données dans ce cas seront établies par le chercheur et la responsabilité de constitution du corpus incombe au chercheur lui-même. Nous voyons donc que l'opération de constitution du corpus demeure importante et non aisée, vu qu'elle détermine les perspectives de l'étude et ses objectifs.

Un certain désir lié à mettre en exergue ce genre de corpus et d'où émanent les problématiques suivantes: Dans quelle mesure et en quoi le traitement de ce genre de corpus oral diffère-t-il? Pour quels objectifs et quels sont les limites ?

**Mots clés :** corpus oral- transcription- analyse de données.

ملخص:

تختلف الدراسات والبحوث العلمية وفقاً لطبيعة العينة اللغوية المراد تحليلها، والتي تحدد أيضاً الطريقة والمحاور الرئيسية للتجربة التي يتم تنفيذها. المساهمة التي نقترحها تطرح عدة أسئلة حول عناصر رئيسية متعلقة بالعينية الشفهية ، تتميز خاصة بمرحلة جمع البيانات خلال عملية النسخ أو التحرير، من ناحية أخرى خصوصية دراسة هذا النوع من البيانات الشفوية.

لذلك من المهم ألا تعتمد عملية التحليل فقط على تسجيل الحدث التواصلي الذي ينتج عن المتحدثين ، بحيث قد يبدو الأمر معقداً للغاية ، لأننا قد نواجه مشكلات تكوين الخطاب المراد تحليله.في الحقيقة ، إنها مجموعة من الخطابات الشفوية المختلفة التي نهدف إلى دراستها. منذ البداية ، نركز على هذا النوع دون غيره، دون لأنه من غير السهل اللجوء إلى جميع الأنواع الفرعية في سياق الأنواع الشفوية حيث أن الباحث في هذا الإطار ملزم بإجراء عملية اختيار بعناية وفقاً لمتطلبات بحثه ، وبالتالي تقديمه وتصنيفه ، للحصول على بيانات متسقة.

نحاول تسليط الضوء على هذا النوع من الخطابات و نطرح بشكل ملح الإشكالات التالية: إلى أي مدى وبأي طريقة تختلف دراسة وتحليل هذا النوع من الخطابات أو العينيات منهجياً عن غيرها؟ ماهي أهداف هذا النوع من الدراسة وما هي الحدود؟

الكلمات المفتاحية : العينة الشفهية – التحرير - تحليل المعطيات

**Abstract:**

Studies and scientific research vary according to the nature of the corpus to be analyzed, which also determines the method and the major axes of experimentation carried out. The contribution we are proposing questions a certain number of major elements concerning the oral corpus, it precisely locates: on the one hand the stage of the data collection in which the transcription operation fits, on the other hand this which is the peculiarity of the treatment of this kind of oral data.

It is therefore important that the process of analysis does not depend solely on the recording of the communicative event produced by the speakers, so that it may seem very complicated, because we are confronted with the problems of corpus formation.

In fact, it is a collection of different so-called oral discourses that we aim to study. From the outset we focus on this genre, without resorting to all subtypes because it is not affordable to start all the specificities of this kind in a single intervention.

Whatever the sketches of such an interdisciplinary study in the context of the oral genres, the researcher is obliged to conduct a selection process carefully according to the requirements of his research, thus of submission and ranking, to obtain consistent data.

The data in this case will be established by the researcher and the responsibility for the constitution of the corpus lies with the researcher himself. We thus see that the operation of constitution of the corpus remains important and not easy, since it determines the perspectives of the study and its objectives.

A certain desire to highlight this type of body and from which emanate the following issues: To what extent and in what way does the treatment of this kind of oral corpus differ? For what goals and what are the limits?

**Key words:** oral corpus- transcription- data analysis.

### **Introduction :**

Dans le contexte de ses tâches quotidiennes simples ou complexes, l'homme a besoin de s'exprimer, le processus de parler ou d'écrire est une forme d'expression , voire de communication pour atteindre différents objectifs et répondre à leurs besoins. Un processus qui repose initialement sur l'utilisation du langage comme moyen véhiculant ses idées.

Il convient en outre de signaler que ce langage est généralement canalisé dans un but distinctif ; déterminé par un contexte particulier. Cela prévoit expressément que l'un des objectifs centraux de ces formes d'expression se devait d'être la communication et l'être avec l'autre, mais aussi la description du monde réel ou imaginaire.

De cette façon, nous stipulons que les deux notions monde et langage sont étroitement liées, de sorte que le processus de communication ou de description ont recours à la trace écrite ou l'événement communicatif produit quelque soit sa forme, ce dernier à son tour apparaît en tant que modèle intégrant les deux mécanismes d'une manière ordonnée régie par des règles non, ce qui rend la définition du corpus au-delà d'une hypothèse d'être uniquement un ensemble de mots ou de phrases, de même s'impose la prise en compte du corpus oral.

Si bien qu'il nous appartiendrait de nous interroger sur la définition de corpus typiquement oral. Le corpus en tant que discours représente un événement communicatif, déterminé par des paramètres de temps, de lieu et de réciprocité ainsi que par toutes les influences non linguistiques dans une

situation donnée. À cet égard, nous considérons que tous les éléments impliqués dans la construction de cet événement communicatif sous le titre de contexte doivent être pris en considération dans le cadre d'analyse discursive.

### 1- Qu'est ce qu'un corpus !

Le terme corpus : « a désigné pendant des siècles des sources documentaires caractérisées par leur exhaustivité: recueils de textes rassemblant exhaustivement tous les documents disponibles pour certains champs d'étude. »<sup>1</sup>. À partir de cette définition, nous notons que le corpus est conçu dans sa conception générale comme un ensemble de données destinées à être analysées selon différents champs d'étude et en fonction de leurs besoins.

En outre, la définition a mis l'accent sur la notion d'exhaustivité, ce qui nous semble contraire à ce que le véritable concept représente, particulièrement quand il s'agit de *données orales ou langues vivantes* dès lors, de quelle exhaustivité parle-t-on !? Cependant, la définition demeure appropriée dans le cas des langues mortes tel que le latin.

Sur cette base, la définition du corpus a pris d'autres dimensions plus larges, d'un point de vue linguistique, il s'agit d'un : « ensemble d'énoncés servant de base à l'analyse »<sup>2</sup> ; cette conception apparaît plus complète que la précédente, car elle ne s'est pas focalisée sur *l'exhaustivité* du corpus ; cela permet de la qualifier pour s'ouvrir ainsi sur d'autres domaines que les sciences du langage à l'image *des sciences humaines et sociales*.

<sup>1</sup>Mireille Bilger, *Corpus: méthodologie et applications linguistiques*, sl, H. Champion, 2000, p.11.

<sup>2</sup> Thomas Hobbes, *De la société civile, de sa matière, de sa forme et de son et de son pouvoir*, Genève, Librairie Droz, 1980, p.137.

En effet, il n'est plus important d'envisager l'exhaustivité en ce qui concerne le corpus, les études reposent sur la collecte des données nécessaires pour configurer un ensemble cohérent, P.Charaudeau signale que le corpus « *n'existe pas en soi* » mais doit être formé par le chercheur étant donné qu'il représente la colonne vertébrale de sa recherche et d'où l'impossibilité de comprendre le corpus sans prendre en compte la recherche qui le développe.

## 2- Mise en point et typologie :

Comme nous l'avons mentionné ci-dessus, un corpus ne peut être un ensemble exhaustif de données, Sinclair dans cette optique le considère comme: « *une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon de langage*<sup>3</sup> ».

Bien que la définition soit claire, elle semble être très restrictive et nous invite fortement à se demander en revenant à ce qui a été énoncé comme caractéristiques, si le corpus oral peut être considéré réellement comme un corpus.

Cela nous a conduit à rechercher une définition qui traite le sujet d'une façon plus profonde :

« *A corpus is any collection of speech recordings which is accessible in computer readable form and which comes with annotation and documentation sufficient to allow re-use of the data in-house, or by people in others organisations.*<sup>4</sup> »

<sup>3</sup> **Claus. Pusch**, *Romance corpus linguistics: corpora and spoken language*, Gunter Tübingen, Narr Verlag, 2002, p.59.

<sup>4</sup> **Delais-Roussarie** Elisabeth , *Corpus et variation en phonologie du français: méthodes et analyses* ,Toulouse, Presses Univ. du Mirail, 2003, p.93.

Autrement dit, un corpus est un ensemble d'enregistrements de parole accessibles sous une forme lisible par ordinateur, et qui contient une annotation et une documentation suffisante pour permettre la réutilisation des données.

En fonction des définitions déjà mentionnées en dessus et en essayant de les analyser, nous présentons dans ce qui suit les types de corpus, en dépit de l'apparente controverse entre les recherches sur la nature de cette notion:

**a- Les corpus écrits :** c'est une collection de textes ou de documents écrits de divers types qui peuvent varier dans leurs caractéristiques, mais demeurent manuscrits tels que 'les œuvres littéraires'.

**b- Les corpus oraux :** comprend des documents et des textes oraux émis instantanément et oralement. L'étude d'un tel corpus exige un recours à la transcription cas : des entretiens, des émissions télévisées... etc.

**c- Les corpus de parole :** collecte d'enregistrements de discours réalisés dans des conditions expérimentales. La nature des données varie considérablement, lecture de mots, lecture de phrases individuelles, lecture de textes...etc. La taille du corpus dans ce cas peut être limitée ou inférieure à celle du corpus oral

**d- Les corpus parallèles :** composés de documents traduits dans une ou plusieurs langues. Par conséquent, certains documents sont la traduction littérale d'autres documents représentant un autre corpus dans la langue source, à partir duquel il a été traduit.

Nous pouvons également concevoir deux autres types de corpus, dans une autre vue qui dépend entièrement de la structure interne et non de la forme ou la structure externe : *corpus de référence* et *corpus de spécialité*.

Dans le premier type (de référence), les documents approuvés pour l'étude offrent la possibilité de découvrir la langue dans son intégralité visant sa

diversité à l'image d'une étude sur (des textes dans la terminologie anglo-saxonne). Par conséquent, la source du corpus varie en fonction de la diversité du mode de communication qui a généré ces textes.

Le second type (de spécialité) est basé sur des données sélectionnées soigneusement pour étudier une dimension spécifique dans un domaine particulier (français médical, une situation de communication ....etc).

### 3- Le corpus oral :

Le discours oral occupe de nos jours une place très importante en sciences du langage, où de nombreux chercheurs ont essayé dans leurs travaux et même en dehors du domaine de la linguistique d'étudier ce type de discours et de mettre en évidence ce genre de corpus : « *les corpus oraux sont des collections ordonnées d'enregistrement de productions linguistique orales multimodales* »<sup>5</sup>  
 Nous déduisons de cette définition simple que le corpus oral est caractérisé initialement par sa multimodalité.

Le but de ce type d'étude dépasse les limites de l'analyse du langage oral notamment le processus d'enregistrement étant devenu utilisé dans plusieurs autres disciplines, tels que les sciences humaines et sociales ou au-delà, en raison de son importance, citant comme exemple : l'étude et d'analyse du discours entre le médecin et le patient durant la consultation afin d'améliorer la pratique de médecine.

### 4- Types de données dans un corpus oral

**4-1-Les données primaires :** « *Constituées par les enregistrements, ayant un lien le plus proche possible avec l'événement documenté* »<sup>6</sup>. Dans ce

<sup>5</sup> **Rastier** François, **Ballabriga** Michel, *Corpus en lettres et sciences sociales: des documents numériques à l'interprétation : actes du XXVIIe colloque d'Albi langages et signification, actes du colloque 2006*, Cedex, cals-cpst, 2007, p.61.

<sup>6</sup> **Sandré** Marion, *Analyser les discours oraux*, Paris, Armand Colin, 2013, p.50.



cas, le corpus peut contenir tous les objets dans le contexte de l'événement. Même les documents écrits ou lus doivent être collectés au cours du processus d'enregistrement des corpus : des images, des objets utilisés, toutes les empreintes mêmes informatiques laissées par l'activité.

**4-2- Les données secondaires :** correspondent au traitement des données primaires elles sont : « *Constituées par la série de descriptions, transcriptions, annotations qui viennent enrichir les données primaires et qui sont souvent fournies après coup et sur la base des données primaires* »

Ainsi, les données dans ce cas recueillent également « *les métadonnées les conventions de transcription et les autorisations des participants*<sup>7</sup> ». En fait, il est vrai que l'ensemble de ces données constitue le corpus oral, mais ce n'est qu'un début de l'étude, parce que plus tard, nous aurons besoin de les organiser et les traiter afin de former le véritable corpus oral.

Dans ce contexte, il s'avère très considérable de noter l'importance des données primaires au cours du processus d'analyse. Car le but d'une telle étude ne se limite pas au processus de transcription du discours, en effet c'est l'étude du discours lui-même.

### **5- Spécificités de l'oral :**

Voici un ensemble de différents points que nous jugeons importants et nous mettons en évidence dans ce qui suit:

---

<sup>7</sup> *ibid.*

-Les données orales sont **matérielles** : « *la notion d'oralité se réfère au versant sonore et physique d'une communication linguistique* »<sup>8</sup> ».

-Il apparaît impossible de faire abstraction de la vocalité (production ou perception d'une voix) pour analyser l'oral.

-L'individualité de la voix qui distingue toute personne d'autrui dite « signature vocale ». Cela interfère intensément dans le processus d'étude du corpus oral car au fur et à mesure que la voix change, nous reconnaissons le changement du locuteur.

- La voix, bien qu'il s'agisse d'une signature individuelle, elle est également liée au groupe et à l'appartenance, où les membres d'une même tribu présentent les mêmes caractéristiques de voix par rapport aux autres tribus.

- Les données orales sont des données globalement **vivantes** produites au cours du processus de perception et de réalisation.

- les données orales sont des données **complexes**, cette complexité se manifeste par le fait que le corpus oral est associé à des caractéristiques non *verbales* (gestes, mimiques...etc) et *paraverbales* : (accent, inflexion, intonation...etc).

L'accent reste mis sur l'impossibilité de prendre en compte tous ces éléments lors du processus d'analyse, conséquemment, nous devons sélectionner le plus important, qui répond aux besoins de l'étude proposée.

## 6- Processus de Transcription :

Quelque soit l'angle sous lequel nous l'envisageons, l'analyse des corpus oraux demeurera une tâche fort compliquée. Marion Sandré stipule que :

---

<sup>8</sup>Siblot Paul , **Bertrand** Véline, **Détrie** Catherine, *Termes et concepts pour l'analyse du discours: une approche praxématique*, Paris, Honoré Champion, , 2001, p.225.

« lorsqu'on a choisi de travailler sur du discours oral, il n'est pas toujours facile de savoir comment enregistrer les données, comment les exploiter, comment interpréter. »<sup>9</sup>

Mondada Lorenza ajoute : « la transcription et l'enregistrement s'éclairent en effet mutuellement : la première permet un accès au second qui en augmente l'intelligibilité et l'analysabilité ; le second donne à la première son caractère d'évidence. Cette relation de réflexivité, dont découle une intelligibilité réciproque, est aujourd'hui matérialisé dans les alignements entre le texte transcrit et le son ou la vidéo »<sup>10</sup>

Autrement dit, l'étude de ce genre de corpus oral porte sur la transcription et dans cette perspective il serait incontestablement désirable de prendre en compte tous les éléments et les données qui se trouvent étroitement liés sous forme d'énoncés oraux.

Ainsi les résultats de l'étude réalisée sur le corpus oral dépendent intensément de ce passage conduisant texte de sa forme orale à sa forme écrite appelé transcription, Bilger voit que : « le travail de transcription n'est pas une question secondaire, au contraire cela suppose et engage toute une réflexion théorique sur les données »<sup>11</sup>. De tout ce qui précède, il convient de rappeler que ce processus de transcription nécessite essentiellement :

-Le respect et la fidélité à l'œuvre d'origine.

<sup>9</sup> SANDRE Marion, *analyser les discours oraux*, Paris, Armand Colin, 2013, p.9.

<sup>10</sup> BILGER Mireille, *Données orales: les enjeux de la transcription*, Perpignan, Presses universitaires de Perpignan, 2008, p79-80.

<sup>11</sup> BARBERIS Jeanne Marie, *Le français parlé: variétés et discours*, Montrelier, Praxiling, Université de Montpellier III, 1999, p.181.

-La prise en considération de toutes les données quelque soit sa forme.

### **6-1- Recourir à la transcription, raisons et stratégies :**

La transcription joue un rôle important qu'il ne faut pas sous-estimer lorsqu'il s'agit du corpus ou discours oral. Des lors, la question qui s'impose est pourquoi transcrire ? En effet Il est toujours utile de rappeler que le rôle la transcription est très important dans les recherches liées au corpus oral, car la réussite de l'analyse et l'atteinte des objectifs de l'étude sur le discours oral sont entièrement liés à la transcription et à la mesure dans laquelle il respecte les critères susmentionnés selon les besoins du chercheur.

Blanche Benveniste évoque dans le même contexte qu' : « *on ne peut pas étudier l'oral par l'oral, on se fiant à la mémoire qu'on en garde. On ne peut pas, sans le secours de la représentation visuelle* »<sup>12</sup>.

En fait, il s'avère difficile d'étudier pleinement le discours oral sans recourir à la transcription pour plusieurs raisons, ce que confirme Traverso : « *La transcription est une préparation indispensable du corpus, à travers laquelle on cherche à conserver à l'écrit le maximum des traits de l'oral* »<sup>13</sup>

### **Pour ce faire, l'opération de transcription nécessite :**

-Une connaissance des deux supports oral et écrit.

-Une connaissance des empêchements ou parfois l'impossibilité de représenter absolument l'oral par l'écrit et que la représentation demeure échelonnée par manque de conditions.

<sup>12</sup>CLAIRE Blanche-Benveniste, *Approches de la langue parlée en français*, Paris, Ophrys, 1997, p.24.

<sup>13</sup> GRECO MORASSO Sara, *Argumentation in Dispute Mediation: A Reasonable Way to Handle Conflict*, Amsterdam, John Benjamins, 2011 , p.156.

-Une connaissance de l'inefficacité de noter tout ce qui a été dit ou vu parce que d'après Traverso : «*Il est impossible, mais aussi inutile de tout noter* ».

### **Conclusion :**

Certes, il est difficile de cerner une telle notion à savoir celle du corpus oral qui échappe à tout emprisonnement intellectuel mais nous avons essayé dans cette réflexion de revisiter la problématique du corpus en particulier oral, nous ne pourrons toujours pas la résoudre en un seul article dont maints points mériteraient d'être mieux dévoilés.

Nous avons essayé un tant soit peu, de consacrer notre recherche à l'étude de cette notion que nous considérons comme très importante au regard de la situation actuelle du monde sous l'influence de l'interdisciplinarité et de différents médias.

### **Bibliographie :**

-**BARBERIS** Jeanne Marie, *Le français parlé: variétés et discours*, Montreuil, Praxiling, Université de Montpellier III, 1999.

-**BILGER** Mireille, *Données orales: les enjeux de la transcription*, Perpignan, Presses universitaires de Perpignan, 2008, p79-80.

-**CLAIRE** Blanche-Benveniste, *Approches de la langue parlée en français*, Paris, Ophrys, 1997.

-**Claus. Pusch**, *Romance corpus linguistics: corpora and spoken language*, Gunter Tübingen, Narr Verlag, 2002.

-**Delais-Roussarie** Elisabeth , *Corpus et variation en phonologie du français: méthodes et analyses* ,Toulouse, Presses Univ. du Mirail, 2003.

-**Mireille** Bilger, *Corpus: méthodologie et applications linguistiques*, sl, H. Champion, 2000

-**Rastier** François, **Ballabriga** Michel, *Corpus en lettres et sciences sociales: des documents numériques à l'interprétation : actes du XXVIIe colloque d'Albi langages et signification, actes du colloque 2006*, Cedex, cals-cpst, 2007.

-**Sandré** Marion, *Analyser les discours oraux* , Paris, Armand Colin, 2013.

-**Siblot** Paul , **Bertrand** Vérine, **Détrie** Catherine, *Termes et concepts pour l'analyse du discours: une approche praxématique*, Paris, Honoré Champion, , 2001.

-**Thomas** Hobbes, *De la société civile, de sa matière, de sa forme et de son et de son pouvoir*, Genève, Librairie Droz, 1980.