

Cours 3 : Définition et introduction au corpus

Introduction

Pour comprendre, analyser et traiter les troubles en orthophonie il est indispensable de disposer d'outils d'observation et d'analyse fiables. Parmi ces outils, le corpus linguistique occupe une place essentielle.

1)- L'importance du corpus en orthophonie :

La formation des orthophonistes exige une solide connaissance des sciences du langage, car la majorité des troubles qu'ils prennent en charge concernent la performance linguistique. Le langage repose sur différents niveaux – phonétique, phonologique, lexical, syntaxique, sémantique et pragmatique – que l'orthophoniste doit être capable d'analyser pour repérer les marqueurs de déviance, à l'oral comme à l'écrit.

La prise en charge implique donc une compréhension fine des mécanismes articulatoires, lexicaux et prosodiques. Par exemple, la rééducation d'une dyslalie (trouble d'articulation) suppose la maîtrise des processus articulatoires, tandis que les pauses anormales dans le discours peuvent révéler des troubles d'accès lexical. Ces analyses exigent une expertise linguistique permettant d'interpréter les faits langagiers à plusieurs niveaux et de proposer des protocoles adaptés.

Voici quelque niveau d'exemples :

Exemple 1 : Phonétique et phonologie

La prise en charge d'un enfant présentant un trouble phonologique (par exemple, la substitution de /ʃ/ par /s/, dire "*sapin*" au lieu de "*chapeau*") suppose de comprendre la différence entre mode et point d'articulation. C'est la connaissance des contrastes phonologiques qui permet à l'orthophoniste de proposer des exercices ciblés pour restaurer l'opposition pertinente.

Exemple 2 : Lexique et accès lexical

Chez un patient aphasique, la production fréquente de paraphrasies sémantiques (dire "*chaise*" au lieu de "*table*") traduit un déficit au niveau du lexique mental. La maîtrise du modèle linguistique de l'accès au mot permet d'élaborer des stratégies de rééducation (renforcement des réseaux sémantiques, exercices de dénomination, catégorisation).

Exemple 3 : Syntaxe

Un patient dysphasique peut produire des phrases agrammaticales ("*moi aller école hier*"). Ce type de trouble révèle une atteinte au traitement morphosyntaxique. L'orthophoniste, grâce à sa connaissance de la grammaire, peut proposer des activités de structuration de phrases et de manipulation syntaxique adaptées.

Exemple 5 : Pragmatique

Chez un enfant avec trouble du spectre autistique, les difficultés à respecter les tours de parole ou à utiliser le langage en contexte révèlent un déficit pragmatique. La connaissance des règles d'usage de la langue permet de mettre en place des jeux de rôle et des médiations favorisant une communication fonctionnelle.

Tous ces exemples montrent que l'orthophoniste mobilise les différents niveaux du langage (phonétique, phonologique, lexical, syntaxique, prosodique, pragmatique) pour analyser les troubles et construire des protocoles adaptés.

Ainsi, l'orthophoniste fonde son intervention sur un double socle : d'une part les données issues de la recherche en sciences du langage, d'autre part le diagnostic individualisé du patient. Sa pratique repose sur une approche globale qui intègre non seulement les paramètres linguistiques, mais aussi des facteurs extralinguistiques (cognitifs, neurologiques, psychologiques, médicaux). Cette interdisciplinarité, qui croise linguistique, sciences cognitives et sciences médicales, rend indispensable la maîtrise du corpus linguistique pour comprendre et traiter efficacement les troubles du langage.

2)- Caractéristique du corpus linguistique :

Un corpus désigne un ensemble structuré d'échantillons de langage (oral ou écrit), recueillis dans des situations naturelles ou contrôlées, et destinés à l'étude scientifique du langage.

En orthophonie, il permet d'objectiver les manifestations langagières, de comparer les productions atypiques aux normes linguistiques et de guider l'élaboration des prises en charge.

A)- Authentique : issu de productions réelles, non artificielles.

Un corpus authentique est constitué de productions linguistiques qui proviennent de situations naturelles de communication. Cela signifie que les données recueillies ne sont pas fabriquées ou inventées par le chercheur, mais qu'elles reflètent la manière dont les individus utilisent réellement la langue dans leur vie quotidienne.

- En orthophonie, cela peut être une conversation spontanée entre un enfant et ses parents, une discussion entre le patient et l'orthophoniste, ou encore la lecture à voix haute d'un texte par un patient dyslexique.
- Contrairement aux exemples artificiels inventés pour illustrer une règle de grammaire ou un test académique, ces productions authentiques montrent les vraies erreurs, les hésitations, les pauses, les reformulations, ou encore les stratégies compensatoires mises en place par le locuteur.
- L'authenticité permet donc de mieux saisir la réalité langagière du patient et de comprendre les troubles dans leur contexte naturel d'apparition.

Exemple concret :

Un enfant dysphasique, lorsqu'on lui demande de raconter une histoire à partir d'images, produit :

« euh... le chat il... tombé... euh... dans l'eau... »

Cette production authentique contient des pauses, des omissions de

morphèmes, une syntaxe simplifiée. C'est ce type de données qui permet à l'orthophoniste d'analyser précisément le trouble, contrairement à une phrase artificielle comme « *Le chat est tombé dans l'eau* », construite par un adulte pour tester la conjugaison.

B)- Structuré : recueilli selon une méthodologie claire (qui, quand, comment, dans quel contexte).

Un corpus structuré est un ensemble de données langagières collectées de manière systématique et organisée, selon des critères préétablis. Il ne s'agit pas simplement d'accumuler des productions, mais de les recueillir en suivant une méthodologie rigoureuse qui garantit leur validité scientifique et leur exploitation clinique.

- **Qui ?** préciser le profil des participants (âge, sexe, niveau scolaire, langue maternelle, type de trouble, etc.).
Exemple : enfants de 7 à 9 ans présentant une dyslexie développementale.
- **Quand ?** indiquer le moment du recueil (phase de rééducation, évaluation initiale, suivi post-thérapie...)
Exemple : enregistrements réalisés lors de la première séance de bilan orthophonique.
- **Comment ?** décrire les modalités de recueil (conversation spontanée, lecture, tâches dirigées, récit à partir d'images, test standardisé, etc.).
Exemple : corpus constitué à partir de récits d'images enregistrés en audio et transcrits en API.
- **Contexte ?** préciser l'environnement (cabinet d'orthophonie, domicile, école), car il influence les productions langagières.
Exemple : interaction parent-enfant enregistrée à la maison pour analyser le langage spontané.

Pourquoi est-ce important ?

Un corpus structuré permet :

- d'assurer la comparabilité entre patients et situations,
- de garantir la fiabilité des analyses linguistiques,

- et de rendre les données réutilisables par d'autres cliniciens ou chercheurs.

Exemple concret :

Au lieu de collecter de façon aléatoire quelques phrases d'un enfant, un corpus structuré consistera à enregistrer 10 minutes de conversation spontanée avec sa mère, dans un contexte familial, puis à transcrire systématiquement les productions pour les analyser (lexique, syntaxe, prosodie, erreurs récurrentes).

C)- Représentatif : il doit refléter la réalité du langage d'un individu ou d'un groupe.

Un corpus représentatif est un corpus qui rend compte fidèlement des usages linguistiques réels d'une personne ou d'une population donnée. Il ne doit pas être biaisé par un recueil trop limité, trop artificiel ou trop orienté vers une seule situation.

- En orthophonie, cela signifie que les productions recueillies doivent refléter la manière dont le patient parle, comprend, lit ou écrit dans sa vie quotidienne, et pas uniquement dans un contexte d'évaluation.
- Pour un individu, la représentativité implique de diversifier les situations : conversation spontanée, récit, tâche de lecture, interaction dialogique. Cela permet de saisir la pluralité de ses compétences (lexicales, syntaxiques, prosodiques...).
- Pour un groupe (par ex. enfants dysphasiques de 6 à 8 ans), il faut constituer un corpus incluant un nombre suffisant de participants, couvrant différents profils, afin d'éviter que les résultats ne soient le reflet d'un seul cas particulier.

Exemple concret en orthophonie :

- Si l'on analyse uniquement la lecture à voix haute d'un texte court, on n'a pas une image représentative du langage d'un enfant dyslexique. On doit aussi inclure ses productions écrites, son discours oral spontané et sa compréhension de textes.
- De même, pour un patient aphasique, il ne suffit pas de noter ses performances lors d'un exercice dirigé : il faut aussi observer

son langage dans une conversation informelle, car c'est là que se manifestent certaines stratégies compensatoires.

D)- Explorable : il peut être analysé quantitativement (fréquence des mots, erreurs) et qualitativement (analyse des structures syntaxiques, organisation discursive, ...).

Un corpus est dit explorable lorsqu'il est organisé de manière à pouvoir être étudié selon différents niveaux d'analyse. Il ne s'agit pas seulement d'un recueil brut de productions langagières, mais d'un matériau structuré, transcrit et annoté qui permet une double exploitation :

1. Analyse quantitative

- Elle porte sur les éléments mesurables du langage : fréquence des mots utilisés, nombre d'erreurs phonologiques, proportion de phrases correctes/incorrectes, temps de pause, longueur moyenne des énoncés...
- Exemple en orthophonie : compter le nombre d'hésitations et de reformulations dans le discours d'un enfant dysphasique afin de mesurer l'ampleur de ses difficultés d'accès lexical.

2. Analyse qualitative

- Elle s'intéresse à la structure et à l'organisation des productions : choix des mots, types d'erreurs (paraphasies, omissions, substitutions), complexité syntaxique, cohérence narrative, organisation discursive, utilisation de la prosodie.
- Exemple en orthophonie : analyser la manière dont un patient aphasique raconte une histoire (enchaînement logique des événements, clarté du discours, utilisation de connecteurs).

Pourquoi c'est essentiel en orthophonie ?

- Le quantitatif permet d'objectiver les troubles (mesurer leur fréquence, leur évolution au fil des séances).

- Le qualitatif permet de comprendre la nature des difficultés et d'orienter la rééducation (ex. : difficulté à organiser le récit ou à choisir le lexique approprié).

3. Types de corpus utilisés en orthophonie

a) Corpus oraux

- Conversations spontanées entre patient et orthophoniste.
- Récits d'images ou de séquences d'histoires.
- Productions de mots isolés (par ex. dénomination).

b) Corpus écrits

- Productions écrites d'enfants (dictées, rédactions, copies).
- Textes d'adolescents ou d'adultes avec trouble acquis (aphasie, dysgraphie).

c) Corpus multimodaux

- Intègrent parole, gestes, prosodie et expressions faciales.

4. Constitution d'un corpus en orthophonie

La constitution d'un corpus suit une démarche méthodologique :

1. **Définition de l'objectif** (évaluer un trouble phonologique ? syntaxique ? narratif ?).
2. **Choix de la population** (enfants normo-parleurs, enfants dysphasiques, adultes aphasiques, bilingues, etc.).
3. **Définition du support de recueil** : tâche libre (conversation), tâche dirigée (récit d'image), épreuves standardisées.
4. **Recueil des données** : enregistrement audio/vidéo ou productions écrites.
5. **Transcription** : en alphabet phonétique international (API) pour l'oral, ou fidèle pour l'écrit.
6. **Annotation** : codage des erreurs, analyse des structures.
7. **Analyse** : quantitative (fréquences, erreurs, longueur moyenne d'énoncés) et qualitative (types d'erreurs, organisation discursive).

5. Applications du corpus linguistique en orthophonie

a) Diagnostic

- Identifier les marqueurs de déviance langagière : erreurs phonologiques (omissions, substitutions), anomalies morphosyntaxiques, troubles lexicaux, incohérences discursives.
- Comparer aux normes développementales.

b) Suivi thérapeutique

- Évaluer l'évolution du patient en comparant différents corpus à des moments distincts.
- Objectiver les progrès (moins d'erreurs, allongement des énoncés, meilleure fluidité).

c) Recherche

- Constituer des bases de données pour comprendre la diversité des troubles (ex. corpus d'aphasies, corpus de dyslexie).
- Établir des normes locales en tenant compte des variétés linguistiques (ex. corpus en français d'Algérie, en kabyle).

Exemple concret

- Cas 1 : Dyslalie
Un corpus de productions orales permet d'analyser les distorsions d'un son [s] en [θ] → sigmatisme interdental. La transcription phonétique précise guide la rééducation.
- Cas 2 : Dyslexie
Un corpus de productions écrites (dictées, lectures) permet d'identifier les régularités des erreurs (confusion b/d, omission de voyelles, erreurs de segmentation).
- Cas 3 : Aphasie
Un corpus de conversation spontanée montre des pauses fréquentes et des paraphasies → indicateurs d'un trouble d'accès lexical.

7. Limites du corpus linguistique

- Nécessite du temps (recueil, transcription, analyse).
- Peut manquer de spontanéité si le patient est intimidé.
- Représentativité parfois limitée (difficile de capter toutes les dimensions du langage en un seul recueil).

Conclusion

Le corpus linguistique constitue un outil fondamental en orthophonie. Il permet d'objectiver les troubles, d'analyser les productions langagières de manière fine et de guider les prises en charge. Sa pertinence repose sur la qualité de la méthodologie et sur l'articulation entre sciences du langage et pratique clinique.

L'orthophoniste se situe ainsi à l'interface entre linguiste (collecte, analyse des données) et clinicien (intervention thérapeutique adaptée).