



HAL
open science

Guide d'annotations collaboratif Corpus Littératie chez les Sourds dit “ Corpus ALPHA ”

Anne Lacheret-Dujour, Caroline Bogliotti, Aurélie Jimenez, Raphaël Prenovec

► To cite this version:

Anne Lacheret-Dujour, Caroline Bogliotti, Aurélie Jimenez, Raphaël Prenovec. Guide d'annotations collaboratif Corpus Littératie chez les Sourds dit “ Corpus ALPHA ”. Guide d'annotations collaboratif du Corpus littératie chez les sourds dit ” Corpus ALPHA”, 2024. hal-04817848

HAL Id: hal-04817848

<https://hal.science/hal-04817848v1>

Submitted on 4 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Guide d'annotations collaboratif

Corpus Littéracie chez les Sourds dit « Corpus ALPHA »

Produit par :

Raphaël Prenovec – Doctorant

Université Paris Nanterre – Laboratoire MoDyCo - ED 139

Dans le cadre de sa thèse de doctorat intitulée « La littératie chez les sourds : états des lieux sociolinguistique et linguistique, perspectives didactiques ».

Ce guide est composé de quatre parties :

1. Partie A : Préambule
2. Partie B : Présentation de l'organisation chronologique du stage et des principes méthodologiques pour l'annotation
3. Partie C : La tâche d'annotation
4. Partie D : Remarques terminales

Partie A. Préambule

L'équipe qui vous accompagnera dans ce travail est composée de Raphaël Prenovec, doctorant, qui a recueilli le corpus, Caroline Bogliotti (MCF Université Paris Nanterre) et Anne Lacheret-Dujour (PR Université Paris Nanterre) qui ont initié la méthode d'annotation, Aimée Lepetit, étudiante en master 2 Psycholinguistique et Linguistique Clinique.

Remarque : parmi les annotatrices qui ont participé à ce travail durant l'année 2020-2021, nous comptons Marion Pombet (étudiante en Master 2), Camille Paulat (étudiante en Master 2), Aurélie Jimenez-Rigal (orthophoniste et étudiante en Master 2), Mathilde Coste (orthophoniste et étudiante en Master 1), Juliette de Montgolfier (orthophoniste et étudiante en master 1) et Esther Gettler (doctorante).

Ce travail de doctorat est financé en partie par l'INJS de Paris (contact : Frédéric Brossier)

Partie B.

Présentation de l'organisation chronologique du stage et des principes méthodologiques pour l'annotation

- **Organisation chronologique**

1. Présentation générale du travail par R. Prenovec : 17 novembre 2021, 24 novembre 2021, 25 novembre 2021
2. Atelier d'entraînement avec l'équipe à Modyco : 6 décembre 2021, janvier 2022
3. Annotation chez soi
4. Brainstorming : séance collective de retour sur le stage mardi 15 mars 2022 :
 - Impressions
 - Difficultés
 - Questions
 - Retours
5. Revue des annotations si nécessaire chez soi après la séance de brainstorming
6. Annotation à finir si besoin
7. Rendu des annotations et du fichier compagnon à R. Prenovec par mail, en copie C. Bogliotti et A. Lacheret-Dujour le 28 février 2022 au plus tard

- **Principes méthodologiques pour l'annotation : fichier excel et fichier compagnon**

Vous allez annoter des erreurs sur un fichier Excel que vous devrez accompagner d'un fichier compagnon au format Word, dans lequel vous allez consigner :

- Erreurs que vous n'arrivez pas à annoter,
- Erreurs repérées par R. Prenovec ou reformulations de RP qui vous semblent discutables
- Toute autre remarque qui selon vous doit faire l'objet d'un retour auprès de l'équipe, par exemple : questions linguistiques intéressantes à discuter collectivement telles que des erreurs qu'on ne retrouverait jamais chez des locuteurs entendants et qui seraient typiques des scripteurs sourds.

Vous présenterez ces points lors de la séance de brainstorming et les décisions à prendre seront prises collectivement.

Chaque fichier Excel contient beaucoup de lignes, souvent, plus d'une centaine, vous serez donc obligés d'annoter en plusieurs séances (pb de fatigue, de concentration, etc.). Il faut penser à consigner pour chaque séance le fichier concerné, les n° de lignes que vous avez annotés (ex : lignes 3-50) et le temps mis à annoter ces lignes dans votre fichier compagnon.

Partie C.

La tâche d'annotation

I. Organisation du fichier d'annotation

Le fichier d'annotation est organisé en 10 colonnes de A à J (l'annotateur ne s'occupe pas des colonnes suivantes) :

I.1. Colonne A et B. Identification corpus

Vous ne vous occupez pas des colonnes A et B qui renseignent l'identifiant du corpus dont sont extraites les lignes à annoter et le genre de discours de ce corpus ;

I.2. Colonne C : Repérage de l'erreur & Colonne D : Reformulation

Vous allez donc intervenir sur :

- Les colonnes C et D préremplies par R. Prenovec ; Les énoncés et les reformulations sont fournis par Raphael Prenovec (RP) dans les colonnes C et D.
- Les colonnes E à J réservées à l'annotation des erreurs proprement dites.

Repérage de l'erreur : l'annotateur doit **noter en rouge** le segment concerné par l'erreur (colonne C) et le segment corrigé correspondant (colonne D) :

Exemple de notations (à gauche : l'erreur > à droite : la correction)

je b_ossais > je **b**rossais
je bossais > je **brosse**

On remarque ici qu'il y a deux erreurs à traiter :

l'omission du « r »

et la terminaison verbale

il doit donc y avoir deux lignes de correction : une erreur par ligne

Soit, en résumé, pour la première étape d'annotation :

1. Vous avez en entrée sur deux colonnes : l'énoncé erreur et sa reformulation ;
2. Vous repérez sur la colonne C en rouge les erreurs et leurs corrections sur la colonne D ;
3. S'il y a plusieurs erreurs dans le même énoncé, RP aura généré plusieurs lignes (une erreur par ligne, cf. fig. 1) ;
4. Vous indiquez **en rouge** l'erreur dans la colonne de gauche et sa correction dans la colonne de droite (cf. fig. 2).

11		Le matin ma famille allent à aeroport charles Gaulle departure vers 11h50 je me suis habille shirt-h et pantalon short il fait vraiment chaud puis l'avion commence ça fait 2h30 dure apres je suis arrive à Minsk au Bielorussie quand je sort l'avion ça fait deja meme quelque second j'ai vraiment froid alors j'ai pris mon veste apres mon ami vient me cherche en voiture Minsk jusqu'a Vitebsk 60km environ puis j'ai rencontre mes amis la-bas c'est vraiment content j'ai amuse bavardé et apres quelque jours je suis allé à Moscou 500km c'est vraiment magnifique la bas c'était super quand je prend le metro et il y avait escalator c'est long apres le metre c'est beau propre.	Un matin, avec ma famille nous sommes allés à l'aéroport Charles de Gaulle pour un départ vers 11h50.
12	0		Je me suis habillé en tee-shirt et en short, car il faisait vraiment chaud. Puis l'avion a décollé et le vol a duré 2h30, et après je suis arrivé à Minsk en Biélorussie.
13	0		Quand je suis sorti de l'avion, ça faisait à peine quelques secondes, j'avais vraiment froid.
14	0		Alors j'ai pris ma veste et après mon ami est venu me chercher en voiture à Minsk pour nous conduire jusqu'à Vitebsk, à 60 km environ.
15	0		Puis, là-bas, j'ai rencontré mes amis.

Figure 1 : exemple de lignes dupliquées par RP

10	0	Tous les matins comme d'habitude je me réveil puis je prend mon petit déjeuner puis je sort et je vais a pied pendant 5 min pour aller au train	Tous les matins, comme d'habitude, je me réveille, je prends mon petit marche pendant 5 min pour aller à la gare.
11	0	Tous les matins comme d'habitude je me réveil puis je prend mon petit déjeuner puis je sort et je vais a pied pendant 5 min pour aller au train	Tous les matins, comme d'habitude, je me réveille, je prends mon petit marche pendant 5 min pour aller à la gare.
12	0	Tous les matins comme d'habitude je me réveil puis je prend mon petit déjeuner puis je sort et je vais a pied pendant 5 min pour aller au train	Tous les matins, comme d'habitude, je me réveille, je prends mon petit marche pendant 5 min pour aller à la gare.
13	0	Tous les matins comme d'habitude je me réveil puis je prend mon petit déjeuner puis je sort et je vais a pied pendant 5 min pour aller au train	Tous les matins, comme d'habitude, je me réveille, je prends mon petit marche pendant 5 min pour aller à la gare.
14	0	Tous les matins comme d'habitude je me réveil puis je prend mon petit déjeuner puis je sort et je vais a pied pendant 5 min pour aller au train	Tous les matins, comme d'habitude, je me réveille, je prends mon petit marche pendant 5 min pour aller à la gare.
15	0	Tous les matins comme d'habitude je me réveil puis je prend mon petit déjeuner puis je sort et je vais a pied pendant 5 min pour aller au train	Tous les matins, comme d'habitude, je me réveille, je prends mon petit marche pendant 5 min pour aller à la gare.
16	0	Tous les matins comme d'habitude je me réveil puis je prend mon petit déjeuner puis je sort et je vais a pied pendant 5 min pour aller au train	Tous les matins, comme d'habitude, je me réveille, je prends mon petit marche pendant 5 min pour aller à la gare.

Figure 2 : visualisation en rouge des erreurs et de leurs corrections

I.3. Quand utiliser le symbole underscore ?

Quand un élément est **absent** alors qu'il devrait être présent, on le signale par l'underscore : « _ » dans la colonne C : Ex : *lai_* qui se réécrit *laid*

Quand un élément est **présent** alors qu'il devrait être absent, on le signale par l'underscore dans la colonne D : Ex : *Demit* qui se réécrit *demi_*

L'underscore est également utilisable pour les mots composés mal écrits où il manque le blanc et éventuellement le trait d'union :

Ex : *biensûr* → *bien_sûr* dans la colonne C et *bien-sûr* dans la colonne D

II. Consignes étiquetage

II.1 Colonnes d'annotation : de E à J (l'annotateur ne s'occupe pas des colonnes K-L-M-N).

Dans le but de d'analyser de manière **quantitative** (statistique) et **qualitative** (linguistique) des erreurs produites, nous avons défini 4 domaines d'erreur, chaque domaine correspondant à une colonne Excel :

- 1) **l'unité linguistique impliquée** dans l'erreur (colonne E. TYPE D'UNITE IMPLIQUEE) ;
- 2) **les propriétés grammaticales de l'unité** (colonne F. PARTIE DU DISCOURS & colonne G. PRECISER LE POS) ;

- 3) le module de traitement (ou composant de la grammaire) auquel appartient l'unité (ex. l'unité syntagme appartient au module syntaxe ; l'unité mot au module lexique, etc.) (colonne H. MODULE DE TRAITEMENT) ;
- 4) l'opération formelle à l'origine de l'erreur (ex. rajout d'un élément, omission d'un élément, substitution d'un élément) (colonne I. OPERATION FORMELLE)

Pour chacun de ces domaines, vous manipulerez un jeu d'étiquettes avec un menu déroulant (figure 3), chaque étiquette représentant l'élément en jeu dans l'erreur. Toutes les étiquettes sont expliquées dans la partie

	E D'UNITE IMPLIQUEE	ARTICLE DU DISCOURS (POS)	Préciser le POS	MODULE DE TRAITEMENT	OPERATION FORMELLE	COMMENTAIRE(S) DE(S) L'ANNO	
1	je prends mon petit déjeuner, je sors et je	Phrase		Ponctuation	Omission	Oubli de la virgule	
2	je prends mon petit déjeuner, je sors et je	Phrase		Ponctuation	Omission	Oubli de la virgule	
3	je prends mon petit déjeuner, je sors et je	Mot		Syntaxe	Substitution	Utilisation du nom "réveil" au lieu du	
4	je prends mon petit déjeuner, je sors et je	Phrase		Ponctuation	Omission	Oubli de la virgule	
5	je prends mon petit déjeuner, je sors et je	Non catégorisable		Syntaxe	Insertion	Ajout de l'adverbe "puis"	
6	je prends mon petit déjeuner, je sors et je	Mot		Morphologie flexionnelle	Substitution	Terminaison en P3 au lieu de P1	
7	je prends mon petit déjeuner, je sors et je	Graphème		Graphisme	Substitution	Oubli de l'accent aigu	
8	je prends mon petit déjeuner, je sors et je	Mot		Phonémique	Omission	Oubli de la lettre "e"	
9	je prends mon petit déjeuner, je sors et je	Phrase		Ponctuation	Omission	Oubli de la virgule	
10	je prends mon petit déjeuner, je sors et je	Non catégorisable		Syntaxe	Insertion	Ajout de l'adverbe "puis"	
11	je prends mon petit déjeuner, je sors et je	Mot		Morphologie flexionnelle	Substitution	Terminaison en P3 au lieu de P1	
12	je prends mon petit déjeuner, je sors et je	Non catégorisable	CONST VERB	marche	Syntaxe	Substitution	Utilisation du segment "vais à pied" au
13	je prends mon petit déjeuner, je sors et je	Proposition	PREP	à	Syntaxe	Substitution	Utilisation de "au" au lieu de "à"
14	je prends mon petit déjeuner, je sors et je	Mot	NOM COMMUN	gare	Lexique	Substitution	Utilisation de "train" au lieu de "gare"
15	je prends mon petit déjeuner, je sors et je	Non catégorisable	PUNCT	point	Ponctuation	Omission	Oubli de point final

Figure 3 : étiquettes avec menu déroulant

Nous décrivons ci-dessous de gauche à droite ces 4 colonnes ainsi que les 2 autres colonnes qui leur sont associées : la colonne « préciser le Pos » et la colonne « commentaires ».

III. 2. Colonne E : type unité impliquée : Étiquettes

On cherche à savoir quelle unité est mal formée dans sa construction interne, du point de vue de l'annotateur non sourd, qui a une intuition sur ce que c'est qu'un écrit bien formé a minima (jugement d'acceptabilité).

La raison pour laquelle un segment est acceptable ou ne l'est pas est une question complexe pour laquelle il n'y a pas toujours qu'une seule réponse, on propose les réponses suivantes en envisageant 6 unités linguistiques qui peuvent faire l'objet d'erreurs, plus une étiquette pour ce qu'on n'arrive pas à classer.

1- Graphème

Le graphème est l'unité graphique correspondant à l'unité orale qu'est le phonème. Un graphème peut correspondre à une lettre, deux lettres (bigraphe) ou trois lettres (trigraphé). Ex : le phonème /o/ dispose de plusieurs graphèmes : o comme dans trop, au comme dans chaud (bigraphe), ou eau dans beauté (trigraphé).

Dans le cas de notre annotation, l'unité touchée est le graphème (la lettre ou l'ensemble de lettres qui représente un phonème donné)

Les erreurs de graphèmes concernent :

- L'omission d'un accent
- le mauvais emploi d'un graphème vocalique ou consonantique
- Écriture SMS

Exemples d'erreur Graphème :
<i>ecrit</i> au lieu de écrit
<i>ékrit</i> au lieu de écrit
<i>écri_</i> au lieu de écrit
<i>tête</i> au lieu de tête .
<i>C ' _ ' tout</i> au lieu de c'est tout

Précision : ne pas tenir compte du point suscrit sur le j (en train de disparaître).

2- Mot

L'unité touchée est l'unité lexicale.

- Soit le mot choisi est transformé par rapport à celui que le locuteur souhaite (ex. ci-dessous avec numéro pour numéro) ;
- Soit une erreur dans le choix du mot, qualifié comme **erreur paradigmatique**, le mot choisi existe mais est inapproprié étant donné le contexte (ex. ci-dessous j'ai mangé un bateau) ;
- Soit une majuscule de nom propre est omise ou une majuscule mise à tort sur un nom commun.

Exemples d'erreur Mot:

Transformation d'un mot cible en pseudomots :

b_o_sé > **brossé**

Nunéro > **numéro**

Substitution par un autre mot du lexique phonologiquement proche mais inapproprié étant donné le contexte

j'ai mangé un bateau > j'ai mangé un **gateau**

Majuscule omise

notre-dame > **Notre-Dame**

3- Syntagme

C'est une erreur dans la construction interne du syntagme : mauvaise combinatoire au sein du syntagme. On doit qualifier cette erreur **d'erreur syntagmatique**. Ne pas confondre avec l'erreur lexicale vue plus haut qui est de nature paradigmatique (mauvais choix sur le paradigme mais la combinatoire est bonne).

Exemples d'erreur Syntagme :

Syntagme mal formé : Construction Verbale touchée > CV correcte est pronominale intransitive

j'habillais de vêtements → je *m'habillais* _.

Mauvaise préposition

Aller à Bretagne → aller *en* Bretagne (pays, région → aller en)

Aller en Nanterre → aller à Nanterre (ville : aller à)

Omission de la flexion de genre

La porte est ouvert ' _ ' → la porte est ouverte

Certaines de ces erreurs s'illustrent en contexte. Ex. pour l'omission de la flexion de genre dans *La porte est ouvert ' _ '* : du point de vue de la morphologie interne, il n'y a pas de faute, car 'ouvert' est parfaitement correct morphologiquement, mais la faute provient du non respect de l'accord à l'intérieur du syntagme.

C'est donc une erreur de combinatoire syntagmatique

4- Proposition

Ici, c'est l'unité **proposition** qui est concernée par une erreur dans sa syntaxe interne.

On devrait surtout rencontrer des erreurs de propositions averbiales (propositions sans verbes), des erreurs de construction des subordonnées (incomplètes, mal enchâssées, etc.).

Exemples d'erreur Proposition:

Manque un pivot verbal :

Ma mère qui arrive_. Elle enlève son manteau. > Ma mère qui arrive enlève son manteau.

Manque un sujet :

J'étais très contente et ' _ ' étais très fière > j'étais très contente et j'étais très fière.

Note : si *j'étais très contente et très fière* est acceptable, *j'étais très contente et étais très fière* ne l'est pas ; le simple fait de répéter le prédicat implique la répétition du sujet également.

5- Phrase

L'unité impliquée est la **phrase** lorsqu'il y a une mauvaise utilisation ou non utilisation des signes de ponctuation (point, virgule, etc mais aussi majuscule omise en début de phrase et substituée par une minuscule). **Attention de ne pas confondre phrase et proposition !** Dans l'erreur phrase, il n'y a pas d'erreur dans la construction interne, ce sont davantage des erreurs de convention graphique.

Exemples d'erreur Phrase :

oubli du point final, oubli de la majuscule, oubli de la virgule :

Ma mère arrive_ elle dine. → *Ma mère arrive. Elle dine.*

C'était du 20 au 23 février 2020. Oui, ça fait déjà au moins une semaine

On refait demi-tour pour regagner la route_ puis on passe devant 2 usines. → *On refait demi-tour pour regagner la route, puis on passe devant 2 usines.*

Paul est parti _ il a oublié sa valise. → *Paul est parti. Il a oublié sa valise.*

Attention, dans *Paul est parti _ il a oublié sa valise*, il n'y a pas deux fautes mais une seule faute de ponctuation qui se traduit simultanément par l'absence de point et la minuscule sur le pronom.

6- Discours

Les erreurs **Discours** concernent les erreurs de sémantique grammaticale et contextuelle, c'est-à-dire le traitement pragmatique.

Les erreurs peuvent être la conséquence d'un **mauvais choix sémantique** d'un mot ou d'un groupe de mots en vertu du contexte sémantique (ex. *aller à la gare* et non *aller au train* ; *marcher pendant 5 minutes* et non *aller à pied pendant 5 minutes* ; si le contexte droit avait été différent, l'expression aurait pu être possible : ex. *aller à pied à la gare* = juste mais *aller à pied pendant 5 minutes à la gare* = faux)

Faute qui se produit au-delà de la proposition (entre deux propositions contiguës ou plus) :

- erreur de cohésion temporelle ;
- problème de connecteurs logiques (cause, conséquence, simultanéité ; etc.)
- chaîne de coréférence (construction de la référence).

Exemple d'erreur Discours

cohésion temporelle

Je me réveille. Je restais dans mon lit pendant 5 min > *Je me réveille. Je reste dans mon lit pendant 5 min*

Utilisation du verbe "rester" à l'imparfait au lieu du présent de l'indicatif.) après la proposition : « je me réveille ».

Erreur : *je me réveille et je restais dans mon lit 5 minutes*

Correction : *je me réveille et je reste dans mon lit pendant 5 minutes.*

les chaînes de coréférence :

Le bébé, ils babillent > *Le bébé, il_ babille*

Attention : ici il y a une seule erreur de coréférence et pas 2, il y aurait eu deux erreurs si on avait eu l'entrée :

Erreur : *Le bébé, ils babille*

(Erreur de coréférence sg-pluriel + erreur d'accord pluriel-sg)

Le garçon, elle vient ce soir > *le garçon, il vient ce soir*

Erreur au niveau de la reprise pronominale, pronom personnel « elle » à la place de « il »).

Note : un verbe conjugué au présent alors qu'il devrait l'être au passé composé ou passé simple par exemple, doit l'être en vertu d'un autre verbe qui le précède pour respecter la cohésion temporelle. Il faut donc qualifier cette erreur comme une erreur fonctionnelle de discours (tel discours fonctionne mal sur le plan syntagmatique), et non pas comme une erreur formelle de morphologie flexionnelle (erreur paradigmatique : mauvais choix dans un paradigme de flexion, la forme appropriée n'est pas acquise).

Une erreur de morphologie flexionnelle concernerait une erreur du type : **vous mangions* au lieu de *vous mangiez*

Chaîne de coréférence : pour les erreurs portant sur l'accord référentiel, vous trouverez probablement plus d'erreurs de nombre (singulier puis pluriel ou vice versa) que de genre)

Pour l'unité Discours, le module de traitement est toujours le module sémantique

7- Non catégorisable

Le segment impliqué est incompréhensible. Aucune étiquette ne convient.

Exemple Erreur Non Catégorisable :

puis on allait à Suisse ce qu'on alla de ponte le plus long on était très fatigüe parce qu'on marche beaucoup envrionne marcher cinq heures par une journée .

→ *ce qu'on alla de ponte le plus long* = Tout ceci doit être considéré comme une seule erreur non catégorisable

II.3. Colonne F : Parties du discours (categories grammaticales des mots)

Le classement que nous proposons est basé sur :

<https://universaldependencies.org/u/pos/all.html>

ADV : adverbe ; la plupart sont des modifieurs de verbes et d'adjectifs

ADJ : adjectif ; modifieur de nom

INTERJ : Interjection = valeur exclamative (ex. Pfff ! waouh !), rentrent ds cette classe également les onopatpées (grrrr !)

NOM : Nom commun ; dénote une personne, un lieu, une pensée, un animal, une idée.

NOM PROPRE : Nom propre

VERB : verbe ; noyau de la phrase

PREP : préposition

AUX : auxiliaire avant verbe prédicatif (*avoir ; être*)

CCONJ : conjonction de coordination (ex. et) à l'intérieur de la proposition

SCONJ : conjonction de subordination (ex. *si, quand*)

DET = Déterminant
NUM : regroupe la classe des déterminants numériques
PRON : Pronom
PUNCT : Ponctuation
SYMB : Symbole
CONNECT : Connecteur ; l'équivalent des conjonctions (de coordination mais au niveau du discours). Typiquement un 'et' ou un 'mais' introducteur de proposition.
CV : Construction verbale

Comment traiter les Constructions Verbales dans =

Exemple : je devrais parler sur l'examen > je devrais parler de l'examen

Type d'unité impliquée = syntagme

Part of Speech- POS = construction verbale

Module de traitement = syntaxe

Exemple : Ce nom est à moi

Type d'unité impliquée = proposition

Part of Speech- POS = construction verbale

Module de traitement = syntaxe

Exemple : j'habillais des vêtements

type d'unité impliquée= syntagme

POS= construction verbale

Module de traitement= syntaxe

Exemple : Aller à Bretagne

Type d'unité impliquée= syntagme

POS= construction verbale

Module de traitement= syntaxe

Exemple : Aller en Nanterre

Type d'unité impliquée= syntagme

POS= construction verbale

Module de traitement= syntaxe

! NON CATEGORISABLE: non traitable. Ne peut être étiqueté.

Normalement, vous ne devriez pas avoir à mettre ce type d'étiquette dans cette colonne

II.5. Colonne G : Préciser le POS

Il faut écrire en toutes lettres l'unité cible, c'est-à-dire l'unité telle qu'elle a été corrigée.

Si ponctuation : il faut indiquer en toutes lettres le signe de ponctuation cible (point, virgule, etc.).

Si majuscule : il faut indiquer la lettre en majuscule.

Exemple POS :

*La LSF c'est mon langage aussi mais il vaut aussi mieux d'apprendre **oraler_***

oraler → Préciser le POS = « **oraliser** »

_ (pour indiquer qu'il manque le point final) → si point Préciser le POS = « **point** »

*m*ontparnasse → Préciser le POS = « **M** »

II.5. Colonne H : Module de traitement (12 modules)

Module (ou composant de la grammaire) qu'on activerait dans un système de génération automatique de texte = boîte à outils qui contient les règles dont on a besoin pour générer un phonème, un graphème, un syntagme, un mot, etc.

Attention : Concernant les erreurs qui touchent l'unité discours, nous avons réfléchi depuis l'atelier du 6/12 et pour des raisons statistiques, nous avons finalement opté pour des règles différentes : voir ci-dessous après la section 11 : Sémantique.

Les modules de traitement sont :

1- Graphisme

Pour toutes les formes graphiques (ex. accent, omission ou ajout de trait d'union ou de blanc dans un mot composé ; majuscule de nom propre).

Exemples Module Traitement Graphisme :

accent **a** à la place de **à** ;

trait d'union

âmes_sœurs au lieu de **âmes-sœurs**, **hand_ball** au lieu de **hand-ball** ;

bien_sûr au lieu de **bien_sûr**.

Majuscule / minuscule

paris au lieu de **P**aris

2- Iconique

Pour toutes formes imagées présentes dans la forme écrite : émoticône, dessin, pictogramme.

Exemples Module Traitement Iconique:

Un smiley dans la phrase *je suis heureuse 😊*

Pour ce type d'occurrence, nous proposons arbitrairement de les considérer comme une erreur d'insertion (on ne met pas (encore) d'émoticones dans les écrits standards).

→ Reformulation = *je suis heureuse _.*

3- Ponctuation

Pour tous les signes de ponctuation, dont la majuscule omise (substituée par une minuscule) en début de phrase

Exemples Module Traitement Ponctuation :

ensuite je prend au moins 7 cuillère a farine = Minuscule à la place d'une majuscule en début de phrase.

Puis on arrive devant un rond point de Guingamp _ il me semble = Oubli de la virgule.

Puis on arrive devant un rond point de Guingamp il me semble _ = Oubli du point marquant la fin de la phrase.

Paul est tombé dans les escaliers. marie a hurlé.

4- Graphémique

Concerne toutes les erreurs de correspondance grapho-phonémique qui **ne génèrent pas de modification sonore de l'unité lexicale**, i.e. si cette unité lexicale était prononcée à l'oral, on ne pourrait pas repérer l'erreur.

Attention : ne pas confondre avec Graphisme (§ 1).

Exemples Module Traitement Graphémique:

Chèse ou chaize au lieu de chaise ;

kartable pour cartable

souphre pour soufre

Le cour_ d'eau pour cours d'eau

un fleuve qui nou_ mene a la plaine pour nous mène ...

je prend au mouin_ 7 cuillère a farine pour moins

<i>Je vais mettre un recipiant dans le four » pour récipient</i>
Note pour le dernier exemple recipiant : le 'e' au lieu de 'é' dans recipiant n'est pas une erreur graphémique ; mais phonémique car ca change la prononciation : /ə/ au lieu de /e/.

5- Phonémique

Concerne toutes les erreurs de correspondance grapho-phonémique qui **gènèrent une modification sonore** de l'unité lexicale

- Le phonème est soit **absent, substitué, inséré, permuté**, et l'erreur phonologique (phonémique ou syllabique) aboutit en principe à un mot qui n'existe pas en Français.
- Problème du hiatus = erreur phonémique par substitution : éviter la contiguïté de deux voyelles en frontière de mot
je habite. au lieu *j'habite* = substitution =>
Si il y a au lieu de *s'il y a*
- S'il y a une erreur d'écriture de l'accent (substitution, insertion ou élision) **et que cette modification entraîne une erreur de prononciation**, alors il s'agit d'une erreur phonémique : *je l'ai rappele* > *je l'ai rappelé* ; *recipiant* > *récipiant*.

Si l'erreur n'entraîne pas une erreur de prononciation (ex è au lieu de ê) alors c'est une erreur graphémique.

Exemples Module Traitement Phonémique :
<i>cratable</i> → <i>cartable</i> = Permutation
<i>cassele</i> → <i>casserole</i> = ici deux omissions : 'r' + 'o', donc deux erreurs
<i>Rejoudre</i> → <i>rejoindre</i> u vs uĕ = Substitution
<i>Sou_enir</i> → <i>souvenir</i> = Omission
<i>Obsevre</i> > <i>observe</i> : phonémique/ Permutation.
Exemples avec contexte :
« <i>devant un fleuve qui nou mene a la plaine</i> » = Oubli de l'accent grave dans "mène".
« <i>on va prendre le fouet et melanger</i> » = Oubli de l'accent aigu dans "mélanger".
« <i>delicatement</i> » = Oubli de l'accent aigu de l'adverbe "délicatement".
« <i>je b_ossais mes cheveux</i> » = Oubli de « r » dans "brossais".

Remarque : l'item cible peut être reconnaissable malgré l'erreur phonologique (*cratable pour cartable*) ou pas du tout (*cassele pour casserole*) dans ce second cas, on reconstitue la plupart du temps l'item cible grâce au contexte phrastique.

6- Lexique

Substitution d'un lexème par un autre (erreur paradigmatique). C'est un mauvais choix dans le lexique mental.

Exemple Module Traitement Lexique :
<i>c'est sur quelle scène</i> pour <i>c'est sur quelle chaîne</i>
<i>Ils m'ont emmenée à l'école</i> pour <i>ils m'ont emmenée au collège</i>

Ce n'est pas une erreur phonémique car le mot choisi est phonologiquement correct (il existe dans la langue). Ces erreurs peuvent être une conséquence de la proximité phonologique avec le mot cible. Ici scène / chaîne sont phonologiquement proches confusion phonémique (/ʃ/ vs. /s/. Le mot peut être remplacé par une unité lexicale non apparentée sémantiquement, apparenté sémantiquement comme dans école / collège.

7- Morphologie dérivationnelle

Mot composé ou mot dérivé mal formé.

Exemples Module Traitement Morphologie dérivationnelle :
Centre_village vs centre ville ;
il nonvalide ma commande vs il invalide ma commande ;
Une chauffeuse vs chauffeure (de bus).

8- Morphologie flexionnelle

La forme appropriée à la personne et au nombre n'est pas celle produite.

- Le suffixe flexionnel n'est pas approprié à la personne et ou au nombre
- L'accord en genre ou nombre entre deux unités n'est pas respecté

Exemples Module Traitement Morphologie Flexionnelle :
<i>je prendrons</i> vs <i>je prendrais</i>
<i>il partirai</i> vs <i>il partira.</i>
<i>Aux</i> chocolat vs. <i>au</i> chocolat
<i>Les amende_</i> vs. <i>les amendes</i>

9- Registre (supprimé)

10- Syntaxe

Les règles de syntaxe doivent être respectées comme :

- la segmentation en propositions complètes ;
- le choix des parties du discours ;
- la construction interne des syntagmes (autres que les erreurs d'accord en genre et en nombre qui seront traités en tant qu'erreurs de morphologie flexionnelle).

Erreur syntagmatique liée à la construction du syntagme ou de la proposition.

Erreur syntaxique (on n'a pas choisi le bon outil, la bonne préposition pour articuler les mots entre eux au sein d'un syntagme ou entre syntagmes à l'intérieur d'une proposition).

Exemples Module Traitement Syntaxe :

On tourne _ un virage > Oubli de la préposition "dans", erreur syntaxique car on dit on tourne **dans** un virage

11- Sémantique

Concerne les règles de cohésion référentielle : la chaîne de coréférence au sein du discours est mal construite ou erronée. Ce type d'erreur est à un niveau plus global que le mot car c'est grâce au contexte que l'on prend conscience de l'erreur.

Ex. « il part » n'est pas faux du point de vue formel, mais si on réfère à « les enfants », alors « il part » est référentiellement faux.

Exemples Module Traitement Sémantique :

Cohésion référentielle (chaîne de coréférence) :

Les enfants, il _ part _ → *Les enfants, ils partent*

12- Cohésion textuelle

Module de traitement impliqué pour les erreurs qui touchent la cohésion textuelle, c'est à dire où l'unité impliquée est le discours mais qui ne sont pas des erreurs de cohésion référentielle.

Deux types d'erreurs sont concernés par ce cas de figure : la **cohésion temporelle** (a dans exemples ci-dessous) et la **cohésion lexicale** (b dans exemples ci-dessous). La première touche l'interface syntactico-sémantique, on signalera donc comme *module de traitement, la syntaxe*. La seconde touche l'interface sémantique grammaticale-sémantique lexicale, on signalera donc comme *module de traitement, le lexique*.

Attention : ne pas confondre syntaxe et morphologie

Illustration : « je me couche » au lieu de « je me suis couché » = module syntaxique, par contre « je me suis couchez » vs. « Je me suis couché » = module morphologique (morphologie flexionnelle)

Exemples Module Traitement Cohésion textuelle : (a) cohésion temporelle

Cohésion temporelle , notamment au niveau verbal :
<i>je me suis lavé les dents et *je me couche → je me suis lavé les dents et je me suis couché</i>
Module de traitement : syntaxe

Exemples Module Traitement Cohésion textuelle : Cohésion lexicale
Le contexte gauche ou droit appelle un mot du lexique, et donc une construction syntagmatique spécifique.
Contexte gauche : <i>aller au train → aller à la gare</i>
CTX droit : <i>aller à pieds pendant 5 minutes → marcher pendant 5 minutes</i>
Module de traitement : lexique

Pourquoi ces choix finalement ?

Car si les colonnes d'annotation sont orthogonales, on peut également les traiter statistiquement de façon interactive. Or si pour toutes les erreurs de discours, on indique la sémantique comme module de traitement, ça ne nous permet pas de dénombrer automatiquement les fautes liées à la cohésion lexicale, les fautes de cohésion temporelle et les fautes relevant de la cohésion syntaxique. Or les différences entre ces 3 types de faute peuvent être significatives, il faut donc un repérage en amont pour leur traitement statistique.

II. 6. Colonne I : Opération Formelle

Vous avez le choix entre 5 opérations :

INSERTION → c'est un ajout de ...

OMISSION → c'est un oubli de ...

SUBSTITUTION → à la place de ...

PERMUTATION → il s'agit d'une mauvaise distribution d'un élément ou d'un segment

TRANSFORMATION → transformé en ...

Exemples d'opérations formelles <u>quand</u> le type d'unité impliquée est le phonème :
Mot cible = cartable
Insertion = <i>caratable</i>
Omission = <i>ca_<u> </u>table</i>

Substitution = <i>castable</i>
Permutation = <i>cratable</i>
Transformation = <i>cabole</i>
Le type d'unité impliquée peut être le mot
Ex pour une omission : <i>je ne mange _ de chocolat</i> pour je ne mange <i>pas</i> de chocolat

Note. Dans certains cas, l'alternative omission/substitution est possible (ex : substitution d'une majuscule par une minuscule ou oubli de la majuscule ; on a opté plus ou moins arbitrairement pour la 'substitution', merci de faire de même.

Partie D. Remarques terminales

Attention aux fausses erreurs :

Ne pas confondre erreur et registre oral familier : on ne note pas comme erreur le registre oral familier mais on peut mettre un commentaire dans la colonne « commentaire » comme par exemple : registre familier inapproprié à l'écrit.

Ex : 'Bah', 'bon'

Aussi : on ne prend pas en compte les phénomènes liés au registre de langue, car c'est trop subjectif. Par contre, on peut faire la remarque dans le commentaire ou dans le fichier compagnon.

Exemple :

Bah ils ne me comprennent pas.

1. Les connecteurs en début de phrase (ex : 'et', 'bref') ne sont pas des erreurs
2. Les symboles numériques ne sont pas des erreurs (ex : 8, 1^{er}, etc.).

Points problématiques pour lesquels vous avez besoin du retour des experts

S'il y a, selon vous, des erreurs dans la reformulation proposée par R. Prenovec, ou des problèmes avec les listes déroulantes : mettre en surbrillance (stabiloter en vert par exemple) la ou les cases concernées, le consigner dans le fichier compagnon et le signaler lors de la séance de brainstorming.

Exemple de problème de typage d'erreurs :

Cas 1 : « Une **épaisse** de forêt ».

Deux interprétations possibles de la forme cible :

(1) « *une épaisse forêt* » vs. (2) « *une espèce de forêt* ».

Si (1) alors erreur syntagmatique : omission de « *de* » qui touche le syntagme donc module Syntaxe ;

Si (2) alors erreur paradigmatique : mauvais choix dans le lexique mental donc module Lexique.

Cas 2 : trop de fautes : trop de transformations, trop de lignes
Surligner en verre, consigner dans le cahier compagnon : fichier, occurrence, ligne

- Attention à ne pas être trop normatif et de faire de l'hypercorrection.

A propos des commentaires : tout commentaire qui, selon vous, pourra être récupéré en aval du traitement pour l'analyse linguistique des erreurs et la compréhension de la spécificité des écrits sourds est particulièrement bien venu.

Exemple :

Cet exemple illustre entre autres l'incapacité de faire de la récursivité en syntaxe en commentaire à l'erreur et à sa reformulation suivante :

Quand j'étais en 7ans,
j'étais en maillot de bain dans les fontaine dans un parc (on a le droit), c'était l'été.