

Enjeux méthodologiques, linguistiques et informatiques pour le traitement du français écrit des sourds



Rapport particulier des sourds à l'écrit⁽¹⁾

- Spécificités linguistiques des productions écrites^(2, 3, 4)
- Niveaux concernés : lexicale, morphosyntaxe, syntaxe

Appels d'urgence 114 FAX – SMS / Problématiques

- faciliter la communication appelants sourds / agents du 114
- Caractériser les phénomènes linguistiques et les types d'erreurs propres au français écrit des sourds
- Confronter écrit sourd à une chaîne de traitement automatique

FAX-ESSU : Exemples choisis

- [1] *je ne vais pas lé drocteur*
- [2] *un voiture est brule*
- [3] *elle est tombé son lit*
- [4] *elle tomber tu lit*
- [5] *tombé sur l'escalier*
- [6] *difficile respiratoire*

Corpus FAX-ESSU

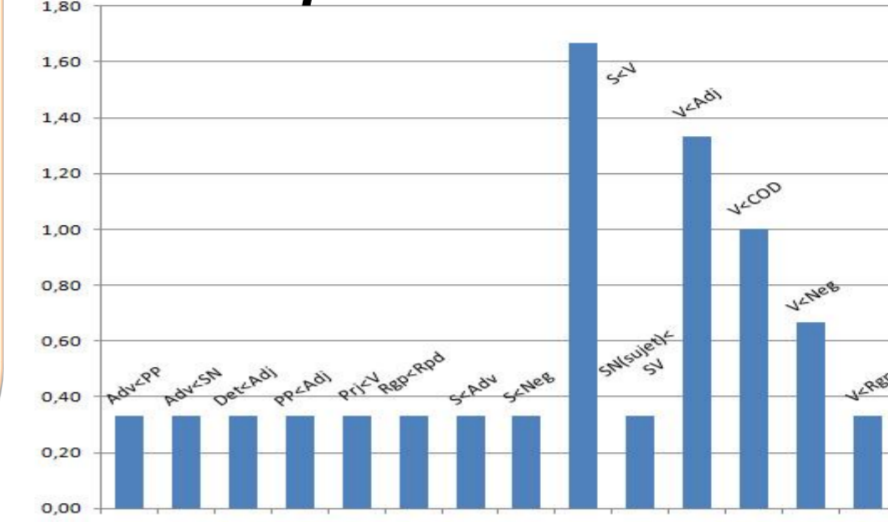
- Corpus élicité de FAX manuscrits en situation d'urgence.
- 300 productions : 20 situations et 15 participants sourds
- 17 mots par production en moyenne.

Méthodologie

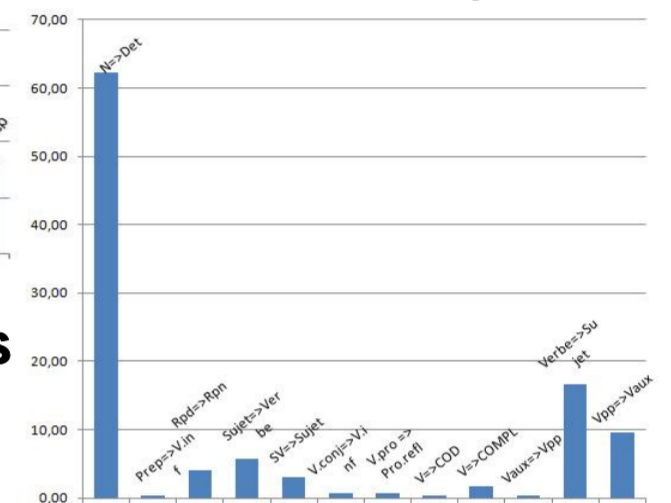
- **Analyses** manuelles et semi-automatiques
- **Niveaux** : lexicale, morphosyntaxe, syntaxe
- **Outils** : Grammaires de Propriétés⁽⁶⁾ et chaîne de traitement LPLSuite⁽⁷⁾.

Syntaxe

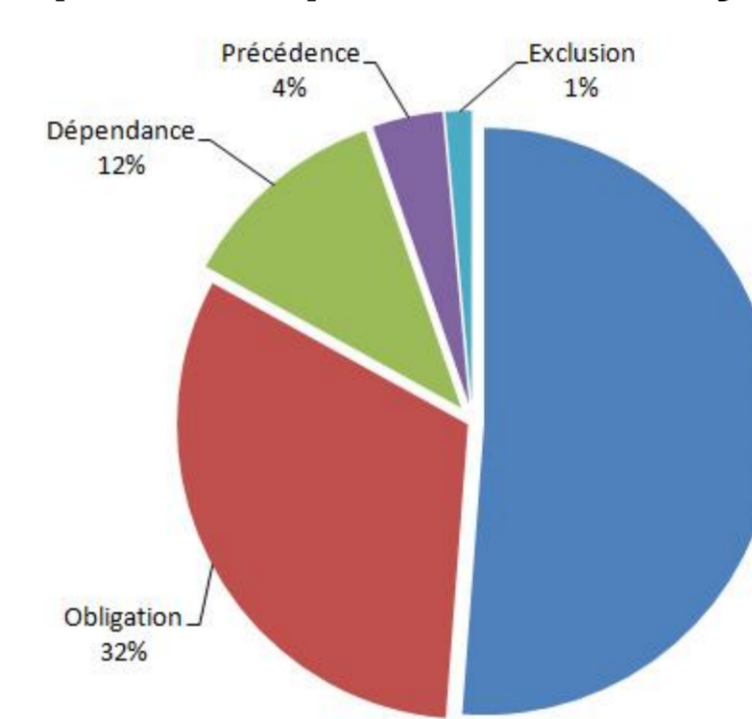
Ordre : *précédence linéaire*



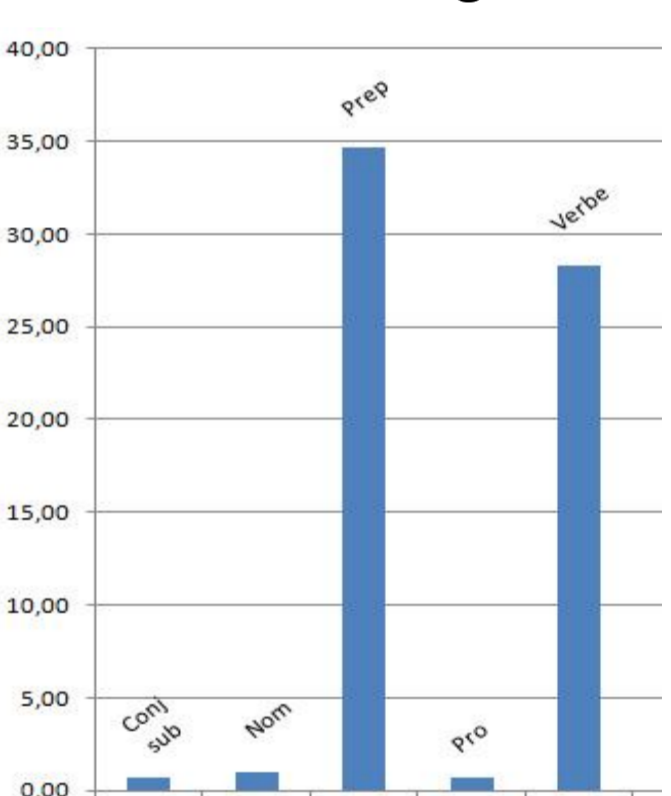
Omission : *exigence*



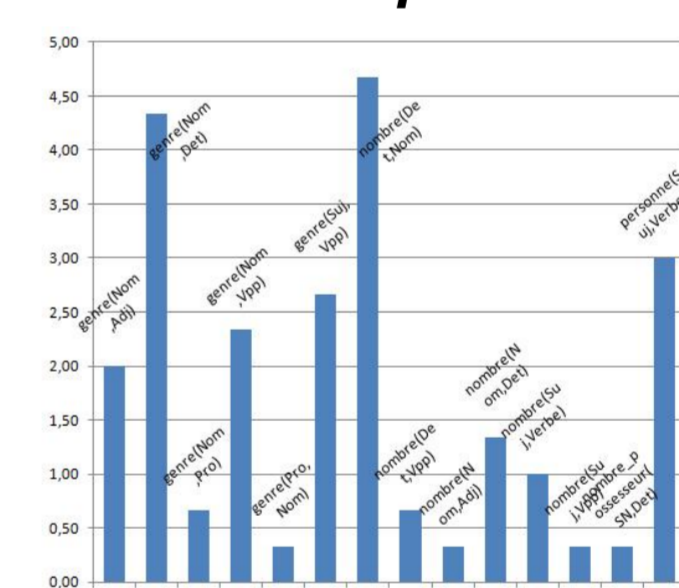
Répartition phénomènes syntaxiques



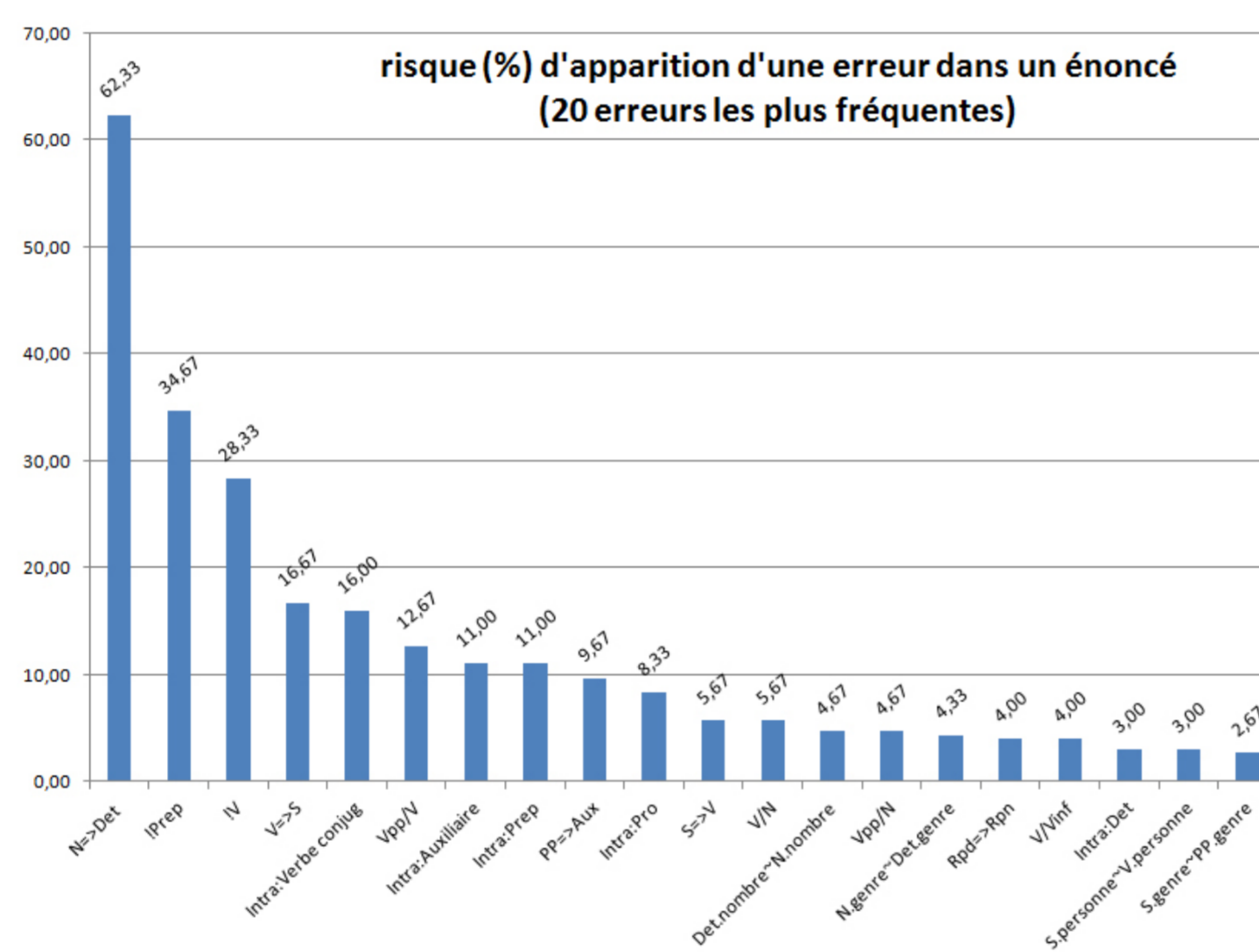
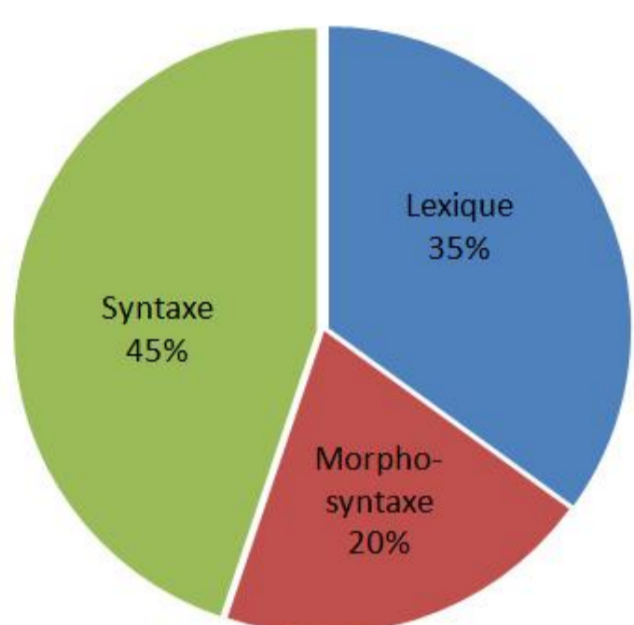
Omission : *obligation*



Accord : *dépendance*



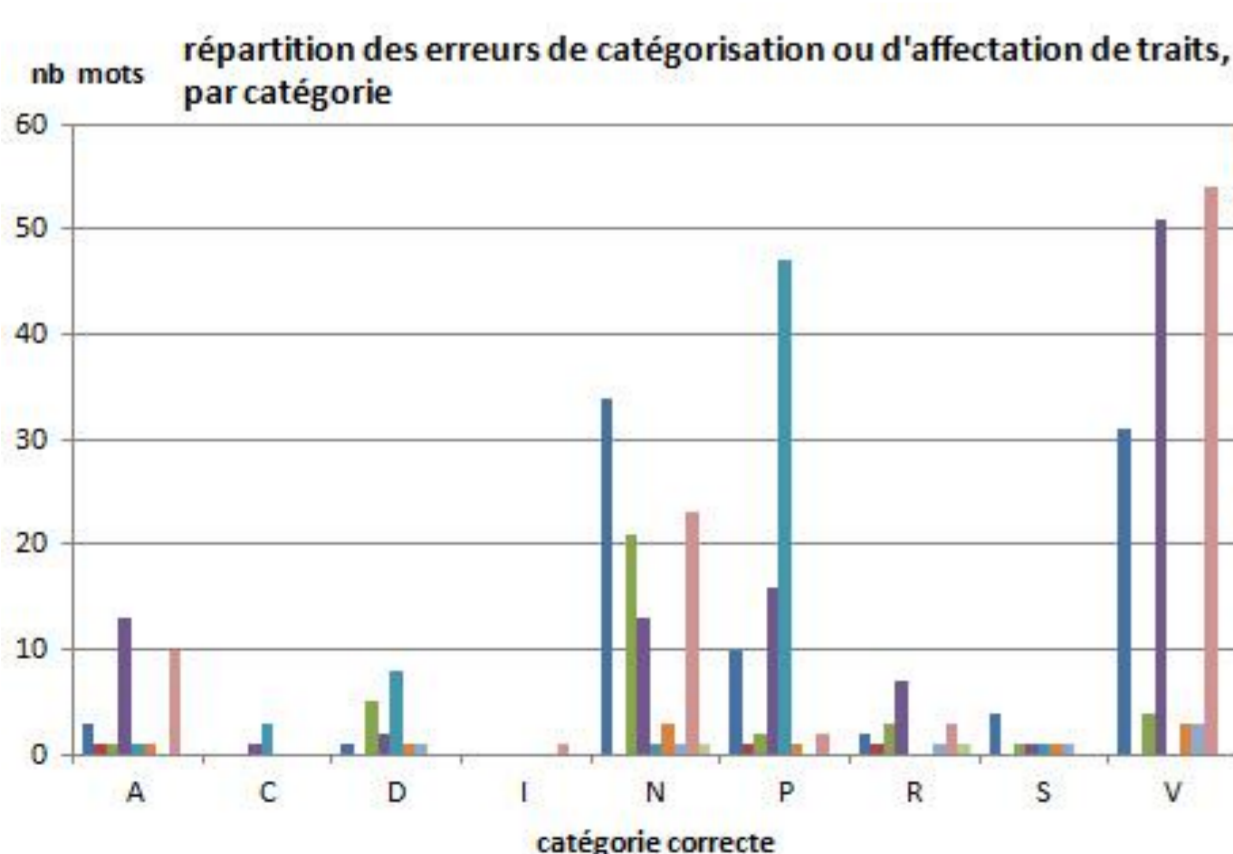
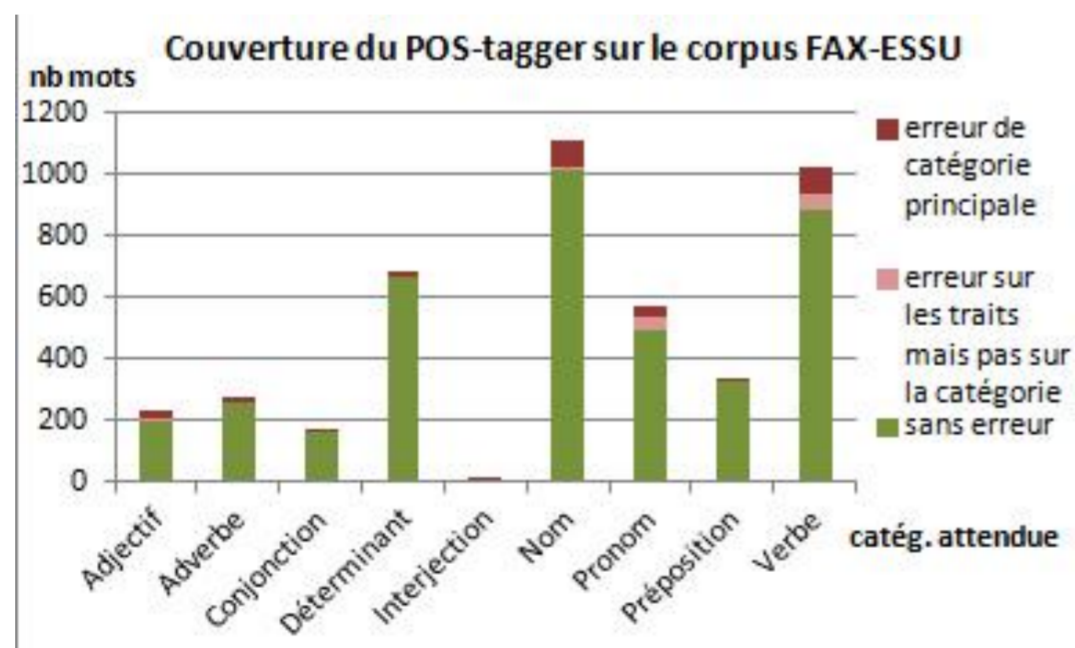
Globalement...



- Syntactique : 25 %
 - Lexical : 35 %
 - Morpho-syntaxique : 20 %
- Recouvrement !

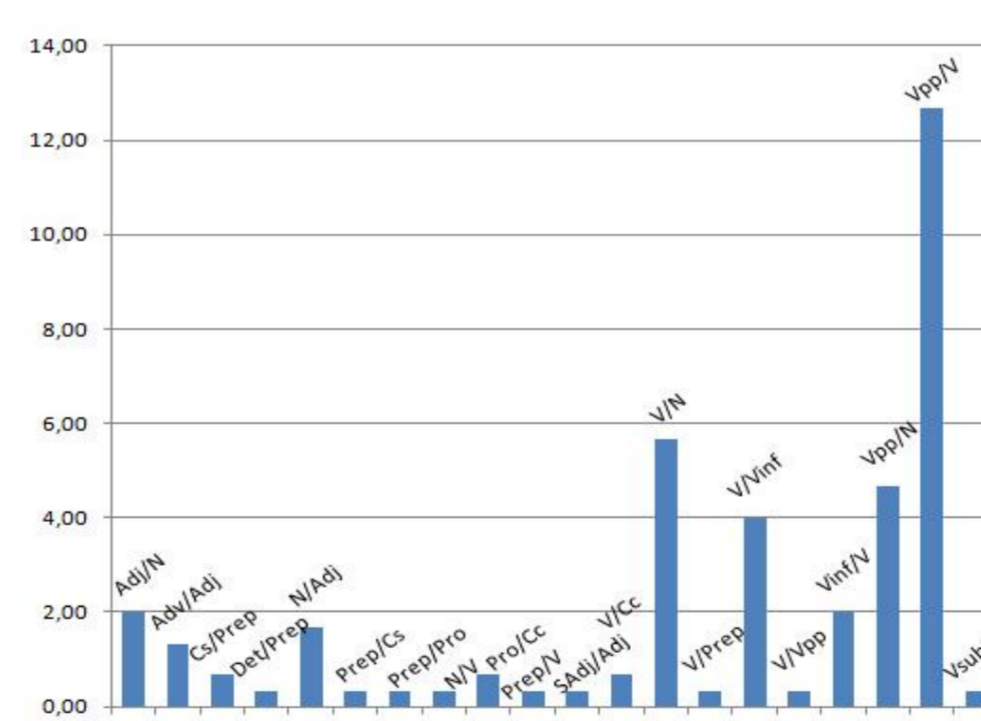
Couverture et efficacité du POS-Tagger

Bonne détection des traits et catégories morpho-syntaxiques...



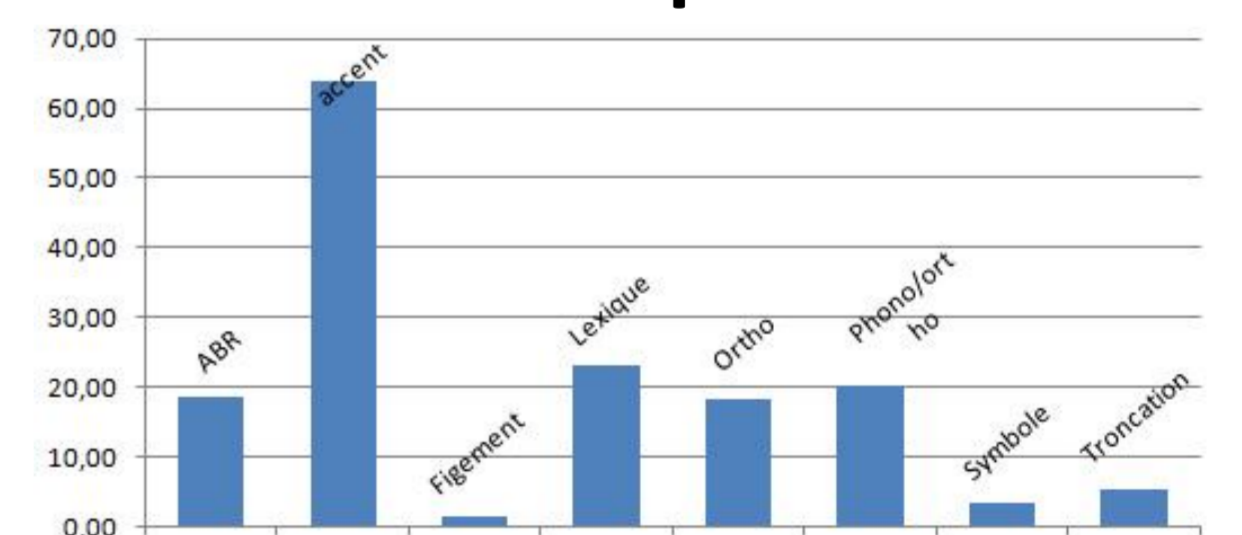
... mais des erreurs liées aux spécificités du corpus

Problème de catégorisation...



- Ou bonne catégorie mais problème de sélection :
- d'un trait grammatical
 - d'un trait sémantique

Lexique



Lexical ou graphique ?
Accent = Vrai problème ?
Quel lien entre ortho et phono ?

Suite et Perspectives

- Affiner ces résultats à tous les niveaux d'analyse =>Linguistique formelle : caractériser les patrons syntaxiques spécifiques au Français Ecrit des Sourds
- Etude et comparaison du SMS 114 avec SMS Sourd⁽⁵⁾ et SMS4Science⁽⁸⁾ : spécificités SMS sourds versus SMS entendants / Mesurer l' impact de la situation d'urgence
- Transfert : formation des agents 114 / pédagogie FLS adapté au public sourd.

Références

- (1) Rapport Gillot 1998
- (2) Nadeau et Machabée 1998
- (3) Périni 2012 / thèse en cours
- (5) Blondel *et al* 2011
- (6) Blache 2000
- (7) Blache, Vanrullen, Balfourier 2006
- (8) Antoniadis *et al* 2011