

Facteurs de variation orthographique dans le grand débat

Enjeux et hypothèse

- Objectif : trouver des facteurs de variation orthographique dans les contributions du grand débat.
- Enjeu : dans un débat, l'orthographe joue un rôle important à deux niveaux : analyses TALN et qualité du débat entre participants.
- Hypothèse : peut-on corrélérer le taux d'erreurs orthographiques à la thématique de la discussion ?

Méthodologie

- ▶ Utilisation d'une liste contenant les 1000 erreurs d'orthographe lexicales les plus fréquentes dans le grand débat.
- ▶ Calcul du taux de variation orthographique dans les 4 fichiers composant la base de données (Démocratie et citoyenneté, Fiscalité et dépenses publiques, Transition écologique, Organisation de l'État et des services publics)

Propositions pour aller plus loin

- ▶ Étendre la liste d'erreurs d'orthographe lexicale et ajouter d'autres types d'erreurs
- ▶ Corréler le taux de variation orthographique à d'autres facteurs :
 - Citoyens vs élus
 - Thématique fine
 - Charge émotionnelle de la contribution (positive, neutre, négative)