

# Frantext catégorisé

## Familiarisation avec le corpus

Sélectionnez « tous les textes » comme corpus de travail. Allez dans « Stats ». Combien cette base textuelle comporte-t-elle de textes, de mots, et sur quelle période ? Quelle est la différence avec Frantext intégral ?

Frantext catégorisé se distingue aussi par la présence d'**annotations métalinguistiques** : les catégories grammaticales des mots (on parle aussi souvent de **parties du discours** – en anglais *part of speech*, POS). Recherchez les flexions du verbe *taire*, et visualisez les résultats. Dans la fenêtre de visualisation, affichez les « Codes ». Quelle information apportent ces codes ? Quelle est leur utilité pour une recherche sur *taire* ?

Recherchez la séquence *en même temps*. Pour Frantext, combien de mots comporte cette locution ?

## Parties du discours

Codes des parties du discours : [http://cid.ens-lyon.fr/aide/ac\\_article.asp?fic=frantext\\_categgram.asp](http://cid.ens-lyon.fr/aide/ac_article.asp?fic=frantext_categgram.asp)

Rechercher une partie du discours : `&e(g=code)`, par exemple `&e(g=V)` pour les verbes.

Recherche un mot en précisant la partie du discours : `&e(g=code c=mot)`, par exemple `&e(g=S c=test)`

Recherche d'un mot fléchi en précisant la partie du discours : `&e(g=code c=&mverbe)`

Quelles sont les onomatopées les plus fréquentes dans la base textuelle ?

Recherchez le verbe *taire* (mot fléchi) au participe passé (partie du discours). Combien d'occurrences trouvez-vous ?

Quel verbe utilise-t-on préférentiellement avec le substantif *enquête* ? Avec *mystère* ? Recherchez la séquence *verbe* suivi de *déterminant* suivi de *enquête* (puis avec *mystère* à la place de *enquête*).