

Exercice Hyperbase

Les marques personnelles dans les vœux présidentiels.

1. Procédure d'intégration du corpus. Mise en œuvre du programme

On utilisera ici la base *vœux*, segmentée et indexée par Hyperbase.

On rappellera qu'il convient d'enregistrer les textes au format txt (texte seul) et de les placer dans le répertoire Hyperbase. La casse originale peut être conservée.

Les partitions permettent de mener des analyses contrastives et définissent les divisions du corpus. Elles sont matérialisées par les caractères distinctifs « &&& », encadrant le nom de chaque sous partie. (Titre d'une œuvre, nom de l'auteur...). La base *vœux* est dotée d'une partition en locuteurs (cinq parties). L'exemple ci-dessous présente l'état du texte balisé avant l'opération de segmentation. (On se reportera au manuel pour des informations plus détaillées)

&&&DeGaulle &&& ... &&&Pompidou&&& ... &&&Giscard&&&
--

Figure 1 : Partition de la base « vœux »

2. Choix du corpus :

Selon la partition préalablement définie, l'utilisateur a la possibilité de choisir les parties du texte ou corpus à soumettre à l'analyse.

→ <i>Boutons de sélection/ Choix du corpus.</i>

Dans le cadre de cet exemple la totalité du corpus sera prise en compte. (Bouton terminer.)

3. Les fonctions documentaires.

- **Biblio** : Cette fonction appelle un fichier qu'on aurait créé lors de la constitution de la base, qui comprendrait toute la méta information du corpus, choix des partitions, informations bibliographiques, conditions de constitution et de saisie ...Ce fichier doit être enregistré dans le répertoire Hyperbase, en texte seul, et comporter l'extension « inf ». (Exemple « vœux.inf »)

- **Lecture** : Permet le retour au texte, au moyen d'une lecture linéaire, dans l'ordre du texte mais aussi hypertextuelle.

(Cliquer sur un mot en affiche la distribution en fréquences absolues, puis les contextes dans la partie souhaitée).

- **Concordance/Contexte** : Recherche d'un mot, affichage en contexte, tris selon l'ordre du texte, contexte gauche ou contexte droit. Les fonctions hypertextuelles permettent à partir d'un mot du concordancier d'en afficher les fréquences, sa position dans l'index hiérarchique...

La recherche peut porter sur la forme graphique mais aussi sur un lemme, une expression, sur une partie du mot (initial, final, chaîne) ou sur une liste de vocables.

- Par lemmes : La recherche porte sur l'ensemble des formes conjuguées d'un verbe à partir de sa forme canonique. Exemple « Vouloir » affichera toutes les occurrences des différentes flexions. Aucune préparation d'étiquetage ou de lemmatisation n'est pré-requise.

- Expression : Syntagme, segment, formule.

- Initial/Final : Recherche à partir d'un préfixe, d'un suffixe.

- Chaîne : Affiche les mots comportant la suite de caractères saisis.

- Tout : Création d'un index, dictionnaire des fréquences et concordances sur l'ensemble du corpus.

- Par liste :

4. Les fonctions statistiques

4.1 Les spécificités

Il s'agit d'un calcul de fréquences probabilisées, permettant d'apprécier la représentation d'un mot dans une partie du corpus par rapport aux autres parties et à l'ensemble (norme endogène) et de porter un diagnostic en termes de sous emplois ou de sur emplois. Le logiciel offre également la possibilité de se référer à un corpus externe (norme exogène).

➔ **Sommaire/Spécificités/ Corpus alphabétique/corpus trié...**

Présentation des résultats :

- Corpus alphabétique : Index alphabétique, écran gauche spécificités positives triées alphabétiquement, écran droit les spécificités négatives (ou déficits), triés de la même façon.

- Corpus trié : Par ordre décroissant de spécificités. (A gauche les spécificités positives, à droite les déficits).

- Tris par parties : DG/Pompidou, Giscard/Mitterrand, Chirac.

➤ **Exemple 1** : Comparer le vocabulaire spécifique de Giscard et Mitterrand.

➔ **Spécificités/tri par partie/Giscard Mitterrand.**

- Sur emploi des marques de la seconde personne du pluriel chez Giscard. Forte représentation d'un lexique lié à la dimension rituelle des messages : « vœux », « souhaite », « bonheur »...

- Mitterrand : Sur emploi très significatif de l'indéfini, importance particulière accordée à l'Europe.

Liens hypertextuels. Là encore, le fait de cliquer sur un mot de ces listes de spécificités en affiche la distribution puis la mise en contexte.

4.2. Graphiques.

A la lumière des listes de spécificités, on peut s'intéresser à une forme particulière et en traduire la distribution statistique sous forme d'histogrammes.

➤ **Exemple 2 :** La forme *Europe* est apparue comme spécifique de Mitterrand, qu'en est-il des autres locuteurs ? Réaliser l'histogramme de la forme *Europe* sur la partition locuteur.

➔ **Sommaire /Graphiques/ Europe.**

➤ **Exemple 3 :** Réaliser l'histogramme des dérivés de la forme *Europe*. (Européen(s), européenne(s), euro...) en utilisant les options de recherche avancée présentées ci-dessus (lemme, forme, chaîne...)

➤ **Exemple 4 :** Les spécificités mettent en lumière des positionnements énonciatifs très différents. (*Vous* chez Giscard, *on* chez Mitterrand...). Réaliser un histogramme double des formes *je* et *nous* sur la partition locuteur.

➔ **Retour/Graphiques/Je/Options/double/nous.**

4.3. Fonction liste

On se propose d'examiner l'ensemble des pronoms personnels et adjectifs possessifs, en recourant à la fonction liste.

On dispose ici des mêmes options de recherche que précédemment (forme, lemme, chaîne, expression...).

- 1. Recherche par forme : On constitue la liste en entrant successivement les formes recherchées.

- 2. Bouton fichier : Permet de charger un fichier qu'on aurait préalablement créé et qui comprendrait la liste des formes à soumettre à l'analyse. (Par exemple l'ensemble des adjectifs possessifs et pronoms personnels).

- 3. Catégories : Il s'agit de listes prédéfinies. Par exemple la catégorie perso123.cls.

Le résultat de ces requêtes se présentent sous forme d'un tableau. Les locuteurs figurent en colonnes les formes recherchées sont présentées en lignes.

- **Exemple 5** : Réaliser la liste des pronoms personnels et adjectifs possessifs selon les possibilités énoncées ci-dessus.

Représentation en histogrammes

En cliquant sur une colonne du tableau (locuteur) on obtient une représentation graphique (histogramme) des spécificités des formes constitutives de la liste pour un locuteur donné. Exemple De Gaulle : Sous emploi des marques de la première du singulier et de la seconde du pluriel, prise en charge de l'énoncé assurée par un « nous » dont le référent est essentiellement la France. (A vérifier par un retour au texte qui peut s'effectuer par une concordance de la première du pluriel chez le premier président de la Cinquième République).

- **Exemple 6** : Réaliser l'histogramme du locuteur De Gaulle, puis de V.G.E.

En cliquant sur une ligne, on obtient un histogramme représentant la répartition d'une forme considérée dans les sous parties du corpus.

- **Exemple 7** : Examiner la distribution du *JE* chez les cinq locuteur du corpus

Représentation factorielle d'une liste.

Hyperbase permet de réaliser des Analyses factorielles, sur des listes de formes. Dans le cas de la liste précédemment établie, on obtiendrait une manière de « panorama énonciatif » des vœux présidentiels.

- **Exemple 8** : Réaliser une analyse factorielle portant sur la liste des marques personnelles.

➔ Lancer AFC/Choisir les deux premiers axes
--

- « Particularité Gaullienne » : Le premier président de la Cinquième République s'oppose sur le premier axe à l'ensemble de ses successeurs. De Gaulle se trouve au voisinage des formes « nous », « lui », « il », « elle ». On observe une proximité de Giscard avec la deuxième du singulier (« toi ») (qu'on explique par de fréquentes apostrophes au temps qui passe « Salut à toi, 1974.... »). Confirmant le diagnostic des spécificités, l'analyse factorielle montre que les marques « eux », « vous », « moi » sont fortement constitutives de l'énonciation Giscardienne. Mitterrand se placerait plutôt du côté du de l'indéfini (*on*), « moi » et de la première personne du singulier.

• • •