

Recueil, constitution, préparation et test d'un corpus lexicométrique (CORDIAL, HYPERBASE, LEXICO, WEBLEX, ALCESTE)

CORRIGE

1. Récupérer dans la base TEXTOPOL les trois débats Bush-Kerry dans la version française parue dans *Le monde* du mois d'octobre 2004.

Solution :

- On peut trouver les trois débats en se reportant aux archives du *Monde* ou dans les CDROM du *Monde*

Observations : il s'agit d'un texte transcrit et traduit à la volée par les soins du *Monde* qui le publie au lendemain des trois débats. Il ne présente aucune garantie de fidélité ni dans la transcription, ni dans la traduction. Une démarche d'expertise nécessiterait l'établissement soigneux du texte américain, sa confrontation à sa version orale audiovisuelle, une analyse sur les données originales, et éventuellement aux transcriptions et aux traductions médiatiques, si l'on s'intéresse plus particulièrement à la transmission de l'information. A défaut de cette démarche le texte ne sera utilisé que comme simple terrain d'exercice lexicométrique. Une version américaine a été également récupérée, dont la fidélité n'est pas mieux établie. Elle pourra être utilisée pour des réflexions sur la traduction en général.

- On trouve aussi le texte des débats transcrit par le Monde dans la base dynamique Textopol

à l'aide de mots « élections, présidentielles, américaines, Bush, Kerry ».

Les textes des trois débats (Floride, Arizona, Washinton) contiennent les interventions des deux candidats, ainsi que celles des journalistes menant les débats et celles des auditeurs dont les questions forment l'armature du deuxième débat.

Plusieurs versions en sont fournies. On retiendra dans un premier temps pour l'expérience une version ne contenant que les interventions des deux candidats regroupées en 6 parties :

B1 : X tours de paroles de Bush dans premier débat

B2 : Y tours de paroles de Bush dans le deuxième débat

B3 : X tours de paroles de Bush dans le troisième débat

K1 : X tours de paroles de Bush dans premier débat

K2 : Y tours de paroles de Bush dans le deuxième débat

K3 : Y tours de paroles de Kerry dans le troisième débat

2. Etablir une liste de questions susceptibles d'éclairer les positions différentes des deux candidats, leurs stratégies, leurs capacités, leurs qualités et leurs défauts.

Eléments de solutions :

On peut rappeler les points préalables suivants :

-L'importance des événements médiatiques constitués par les trois débats dans l'agenda de la campagne, présentés sinon comme cruciaux, du moins comme des moments importants dans la course au pouvoir démocratique

- Le candidat Kerry a été donné assez nettement vainqueur des trois débats par une majorité des médias et des commentateurs au lendemain de chacune des confrontations

-Bush a remporté néanmoins une victoire électorale.

Il ne s'agit bien sûr pas de vouloir expliquer rétrospectivement par la statistique lexicale, la victoire ou la défaite finale des candidats. On peut tout au plus essayer **d'éclairer** les faits contradictoires à partir d'observations et de constats lexicaux, grammaticaux, stylistico-rhétorique, en faisant abstraction de tous les phénomènes interactionnels, verbaux, gestuels, comportementaux, représentationnels constituant les confrontations médiatiques.

On peut établir par exemple des constats concernant

-Richesse du vocabulaire

-Nature du vocabulaire

- Thématiques dominantes

-Vocabulaire partagé

-Vocabulaire spécifique

-Catégories grammaticales spécifiques

-Schéma syntaxique spécifique

3. Répondre à ces questions à l'aide de CORDIAL, (caractériser les formes syntaxiques principales et les thèmes globaux du corpus, etc.)

Un passage dans CORDIAL permet d'étiqueter un texte et de quantifier un nombre énorme de variables (plus de 1000) un les dénombrant, en en donnant la répartition relative dans un texte, ou en comparant ces résultats à un corpus extérieur.

On a constitué 6 textes

4. Constituer un corpus au format HYPERBASE, un corpus au format LEXICO3, un corpus au format ALCESTE , un corpus texte brut (objet de l'exercice 1.2)

A. Observations initiales à partir du logiciel HYPERBASE

On peut décrire d'abord les caractéristiques statistiques du corpus BK6 à l'aide du menu structure

1. Le **Bouton structure** donne l'étendue du vocabulaire particulier à chaque partie et du vocabulaire partagé entre elles

1.1. Le **Bouton étendue** donne la longueur des 6 parties et l'étendue de leur vocabulaire; représentables par un histogramme

1.2. Le **Bouton Hapax et richesse** : donne une évaluation de la richesse et des mots rares; représentables par un histogramme

1.3. Le **Bouton Hautes fréquences** donne les 200 mots les plus fréquents du corpus global. On peut chercher les mots les plus fréquents de chaque partie et commenter la répartition des noms, des verbes, des mots grammaticaux. On peut alors étudier la répartition de certains mots qui paraissent intéressants, en allant dans le menu Mots et contexte

1.4 Le **Bouton distance** permet de représenter par une AFC ou par arborexence la proximité des différentes parties

2. Le **Bouton Spécificités** permet alors d'examiner les listes des mots les plus caractéristiques de chaque partie du corpus, de les commenter ou de les analyser à l'aide des bouton mots et concordance ou contexte

3. Le **Bouton Liste** permet de constituer des listes, ou ensemble de formes, intéressantes (les pronoms, les verbes modaux, les mots les plus fréquents, les termes désignant le terrorisme, la guerre, la sécurité, la religion, la famille etc) et d'établir des AFC pour ces listes, ou d'en analyser la répartition par histogramme.

On peut ensuite systématiser et approfondir les descriptions dans les cadres suivants.

B. Distributions statistiques (Tableau Lexical Entier)

1. Propriétés zipféennes du TLC

Les paramètres initiaux

Le rapport rang-fréquence

La richesse lexicale

Les hapax

Les THF

2. Sur le vocabulaire

Voc partagé

Voc original

Voc spécifique

3. Proximités intertextuelles

AFC

CHA

C. Distributions linguistiques (Tableaux Lexicaux Réduits)

1. Catégories énonciatives

Pronoms et déterminants du discours

Modalités : assertives, interrogatives, exclamatives, jussives

Négations

Les auxiliaires modaux : Vouloir, pouvoir, devoir, falloir, avoir à

2. Catégories grammaticales

Noms et verbes

Adjectif et adverbe

3. Approches lexico-thématiques

5. Répondre à ces questions à l'aide d'HYPERBASE, (contraster les partitions du corpus en 6)

6. Répondre à ces questions à l'aide de LEXICO3, (contraster diverses partitions du corpus)

7. Répondre à ces questions à l'aide d'ALCESTE, (Rechercher des partitions dans le corpus)

8. Etablir une synthèse et une comparaison des diverses approches.