

Conversion de balisage Lexico vers Alceste

Bush - Kerry trois débats traduction française

Exercice 5.1 Exercice de balisage sur le corpus Bush Kerry

Le corpus :

Fichier : 3Debfr.rtf, (base dynamique) Trois débats Bush Kerry en Français, transcription *Le Monde*.
Casse originale
Encodage Lexico. Partition en :

- Débats (<DEB=1>, <DEB=2>, <DEB=3>)
- Candidats (<CAND=B>, <CAND=K>)
- Locuteurs (intervenants extérieurs conservés) (<LOC=B1>, <LOC=B2>, <LOC=B3>, <LOC=K1>, <LOC=K2>, <LOC=K3>, <LOC=Q1>, <LOC=Q2>, <LOC=Q3>)

Convertir le balisage Lexico en mots étoilés de façon à obtenir une configuration conforme à l'exemple suivant :

```
**** *DEB_1 *LOC_K1 *CAND_K  
□ désignant l'espace
```

1. Matérialiser les Unités de Contexte Initiales (U.C.I).

- ✓ **Procédure** : Rechercher le motif <DEB= et remplacer par **** *DEB_
Puis **Remplacer tout** (ne pas oublier l'espace avant le premier mot étoilé)

2. Distinguer les locuteurs des intervenants extérieurs

On remplacera, seulement pour les questions, la balise « LOC » par une balise « INT »

- ✓ **Procédure** : Rechercher <LOC=Q1> remplacer par <INT=Q1>, <LOC=Q2> par <INT=Q2>, <LOC=Q3> par <INT=Q3>
A chaque étape : **Remplacer tout**

3. Transformer la balise Candidats (CAND) en mot étoilé et supprimer le saut de ligne :

- ✓ **Procédure** : Rechercher le motif >^p<CAND=
(^p désigne le passage à la ligne)
Remplacer par CAND_ (ce motif sera précédé d'un espace)
Puis **remplacer tout**

4. Supprimer les balises > < de la première ligne de l'UCI

- ✓ **Procédure** : Rechercher le motif ><LOC= et remplacer par *LOC_ (précédé d'un espace)
Puis **remplacer tout**

5. Supprimer les balises en fin de ligne

- ✓ **Procédure** : Rechercher >^p et remplacer par ^p
Remplacer tout

6. Traiter le balisage des interventions extérieures

- ✓ **Procédure** : Rechercher ><INT= et remplacer par *INT_ (motif précédé d'un espace)

7. Modification de la casse, convertir le texte en minuscules

8. Enregistrement du fichier

Format texte seul (txt avec sauts de lignes) dans le répertoire « Analyses Alceste »

Pour finir...

Lancer Alceste, créer un nouveau plan d'analyse et vérifier le formatage du corpus.
« Il n'y a pas d'erreurs de saisie » !!

Vers l'exercice 5.2...

- Le rapport d'analyse
- Description des classes
- UCE classées, rejetées
- Tri croisé sur une forme
 - Terrorisme ?
 - Président ?
 - Je, nous ?
- Tri croisé sur une variable
 - CAND ?
 - LOC ?
 - DEB ?
- Extraction, exportation des mots étoilés

• • •