

## **Projet DILAF**

### **Dictionnaires Langue Africaine – Français**



## **Méthodologie DILAF**

### **Étape 5 : Contrôle des listes fermées**

#### **Objectif**

Standardiser les noms des catégories lexicales, des classes, etc.

#### **Outil**

Tableur Open Office

Exchanger XML Editor

#### **Motivation**

Certaines informations figurant dans un dictionnaire appartiennent à une liste fermée. C'est le cas par exemple, des étiquettes lexicales, des classes, ou encore des indications de genres.

Ces informations souffrent parfois d'un manque de normalisation : des abréviations varient, il y a des erreurs, etc.

Il est nécessaire de dresser ces listes d'informations utilisées dans le dictionnaire puis de les examiner et d'effectuer les éventuelles corrections.

## Méthode

### 5.1 – Isoler les informations à contrôler

Le dictionnaire est copié dans un fichier temporaire. Toutes les informations (sauf celles à contrôler) sont ôtées. Cette opération est facile car les informations sont toutes repérées et encadrées par des balises.

La dernière version du dictionnaire (hausa\_extrait\_v72.xml) est sauvegardée sous le nom `etiquettes.xml`

Pour obtenir les classes utilisées dans le dictionnaire, il suffit d'ôter toutes les informations sauf les étiquettes.

La DTD, l'entête XML, les balises dilaf (ouvrante et fermante) sont ôtées.

Le dictionnaire est sauvegardé sous le nom `etiquettes_v2.xml`

Les informations autres que les classes sont ôtées à l'aide d'une expression rationnelle (voir transformations .odt)

Le dictionnaire est sauvegardé sous le nom `etiquettes_v3.xml`

Les notices restantes, qui n'avaient pas de classes, sont ôtées autres à l'aide d'une expression rationnelle :

Le dictionnaire est sauvegardé sous le nom `etiquettes_v4.xml`

### 5.2 – Contrôler

Les informations qui ont été isolées doivent être contrôlées. Or, leur liste peut être longue (plusieurs milliers d'items). Pour les contrôler efficacement il est utile de les regrouper afin de les scruter plus facilement. cette opération peut être réalisée à l'aide de la fonction de tri d'un tableur.

Le contenu du fichier peut être alors copié dans un tableur en faisant un copié-collé de tout son contenu.

Le tableur est sauvegardé sous le nom `etiquettes_v5.ods`

Le tableur est trié.

Le tableur est sauvegardé sous le nom `etiquettes_v6.ods`

La liste des étiquettes utilisées dans le dictionnaire peut alors être précisément dressée.

### 5.3 – Corriger

La liste des informations à scruter étant dressée, il est facile de la parcourir afin de détecter des anomalies (par exemple une classe notée "s" au lieu de "s."). Il faut alors se reporter à la dernière version du dictionnaire et effectuer les corrections nécessaires dans celui-ci (par exemple, les classes notées "s" sont corrigées en "s.").

Dans notre dictionnaire, les classes sont indiquées de manière cohérente et homogène pour tout le dictionnaire. Il n'est donc pas nécessaire d'effectuer des corrections.

Une démarche identique doit être effectuée pour vérifier la liste des genres.