

Travaux Dirigés de linguistique

Ressources linguistiques

—Master 1—

Dictionnaires

Nous allons étudier dans cette fiche les dictionnaires électroniques de mots simples.
Nous verrons les méthodes de flexion automatique et de compression de dictionnaires.
Les dictionnaires DELA sont détaillés dans le chapitre 3.1 du manuel.

► Exercice 1. Flexion automatique

1. Construire des transducteurs permettant de fléchir les mots suivants : *bocal*, *cheval*, *cochon*, *instituteur*, *oeil* et *pain*.
2. Construire des transducteurs permettant de conjuguer au présent les verbes *acheter* et *peler*.
3. Fusionner les 2 transducteurs précédents en un seul grâce à l'opérateur R.
4. Donner le DELAS correspondant aux mots précédents. Attention, *cochon* peut être un nom ou un adjectif.
5. Donner le DELAF obtenu par la flexion du DELAS précédent.

► Exercice 2. Compression

Afin de représenter un dictionnaire de façon compacte, nous séparons chaque entrée en deux éléments :

- la forme fléchie ;
- une étiquette comprimée permettant de reconstruire l'entrée complète à partir de la forme fléchie.

Pour reconstruire l'entrée complète, on doit pouvoir retrouver la forme canonique à partir de la forme fléchie. Pour cela, on code le nombre de lettres à enlever, suivi du suffixe à ajouter. On concatène ensuite la séquence obtenue avec les informations grammaticales et flexionnelles.

Exemples :

`jeux, jeu.N:mp` a pour étiquette comprimée `1.N:mp` (enlever 1 lettre)

`chantons, chanter.V:P1p` a pour étiquette comprimée `3er.V:P1p` (enlever 3 lettres et ajouter er)

`vraiment, vraiment.ADV` a pour étiquette comprimée `.ADV` (ne rien enlever ou ajouter)

Une fois que l'on dispose de ces étiquettes comprimées, il suffit d'associer à chaque forme fléchie ses différentes étiquettes comprimées. Cette association sera faite à l'aide d'un transducteur que l'on minimisera. La minimisation permettra de compresser le dictionnaire, tout en accélérant l'accès à ses entrées.

1. Associer à chaque ligne du DELAF de l'exercice précédent son étiquette comprimée.
2. Associer à chaque forme fléchie la liste des étiquettes comprimées avec lesquelles elle peut se combiner.
3. En numérotant les listes ainsi construites, construire un transducteur minimal associant à chaque forme fléchie le numéro de la liste qui lui correspond.

► **Exercice 3. Flexion et compression d'un dictionnaire sous Unitex**

1. Construire des transducteurs permettant de fléchir les mots suivants : *acheter*, *bocal*, *cheval*, *cochon*, *instituteur*, *oeil*, *peler* et *pain*. Les verbes *acheter* et *peler* ne seront conjugués qu'au présent de l'indicatif. Sauver ces grammaires dans votre répertoire `/unitex/French/Inflection`.
2. Compiler les transducteurs obtenus ("Compile FST2" dans le menu "FSGraph/Tools").
3. Créer le DELAS correspondant aux mots précédents dans un éditeur de texte et sauver le fichier sous `/unitex/French/DELA/delac.dic`.
4. Utiliser les transducteurs précédents pour fléchir automatiquement le dictionnaire `delas.dic`. Pour cela, ouvrir le fichier avec "Open" dans le menu "DELA". Lancer la flexion en cliquant sur "Inflect...".
5. Ouvrir le DELAF obtenu (`delasflx.dic`) et vérifier son format ("Check Format...").
6. Compresser ce DELAF en cliquant sur "Compress into FST". Lancer la commande `more` sur le fichier `delasflx.inf`.
7. Appliquer ce dictionnaire ("Apply Lexical Resources...") lors du chargement du texte *Le tour du monde en 80 jours* comme vu au TP précédent. Comparer avec l'application du dictionnaire par défaut.
8. Construire l'automate du texte ("Construct FST-Text..."). Quelle conclusion peut-on en tirer ?