

U.E. Informatique pour le TAL (SL02244X)

Manipulation de données XML en Perl

Franck Sajous - CLLE-ERSS
`sajous@univ-tlse2.fr`

29 mars 2012

Ce document est disponible à l'adresse :

<http://w3.erss.univ-tlse2.fr/membre/fsajous/SDL/SL02244X/>

1 Corpus *AirFrance*

Extraire du corpus *AirFrance* les tours de parole contenant des pauses, en indiquant pour chaque tour de parole de quel locuteur il s'agit. Le fichier texte produit en sortie devra avoir la forme suivante :

[C] : en fin de matinée [PAUSE] ou entre midi et deux
[0] : 2020 francs certainement attendez je le vérifie [PAUSE] mais ça me parait tout à fait plausible

2 Lexique Morphalou

2.1 Extraction d'un lexique dérivé formes fléchies/lemmes

À partir du lexique Morphalou (on utilisera la version 1.0), extraire, pour chaque forme fléchie, la liste des lemmes possibles. Le fichier texte produit en sortie devra avoir la forme suivante :

dromadaire	dromadaire
dromadaires	dromadaire
souris	souris sourire

2.2 Conjugueur

Écrire un programme qui demande à l'utilisateur de saisir la forme d'un verbe à l'infinitif et qui conjugue ce verbe (on se limitera au présent de l'indicatif et au futur simple).