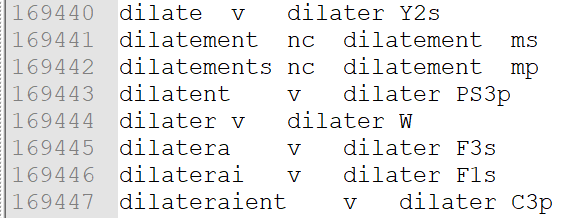
**Rapport sur la construction du dictionnaire du français**

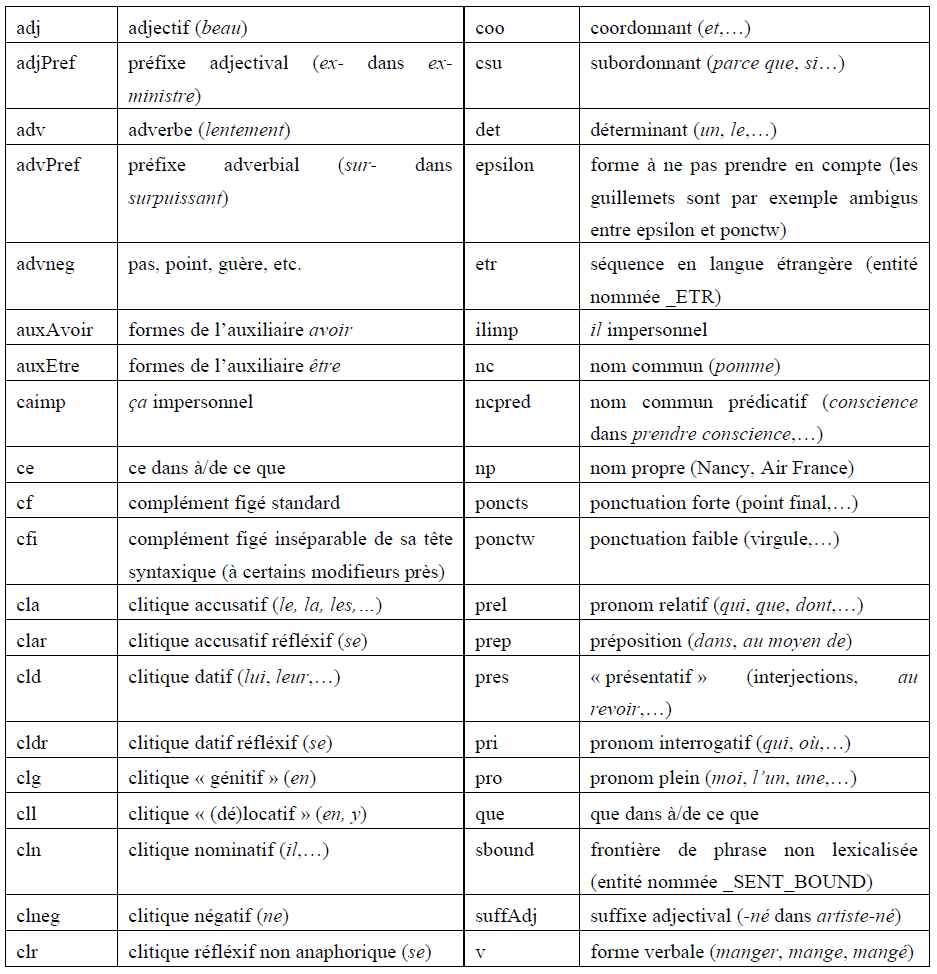
Pour l’instant j’ai construit un dictionnaire qui contient 546432 mots à la base du fichier *lefff-3.4.mlex*.

Je vous explique mes procédures de construction.

1. Compréhension du lefff :

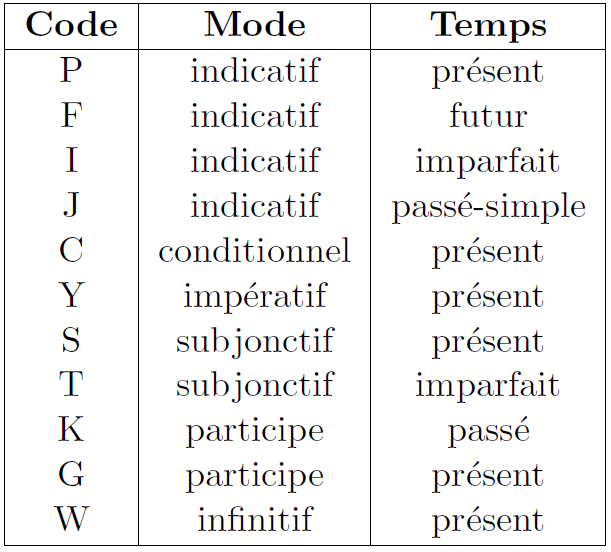


* Dans *lefff-3.4.mlex,* la première colonne indique le mot, la deuxième colonne est la catégorie, la troisième colonne est le lemme, la quatrième colonne est le trait.
* Jeu de categories utilisées dans le Lefff :

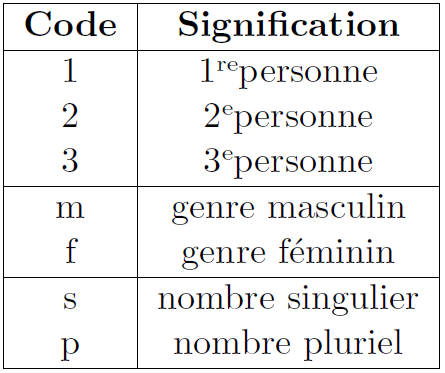


* Codes pour les traits :

- Un code indiquant le mode et le temps : P, F, I, J, T, Y, Z, S, C, K, G, W.



- Suivi d'un ensemble de lettres représentant la personne, le genre et le nombre.



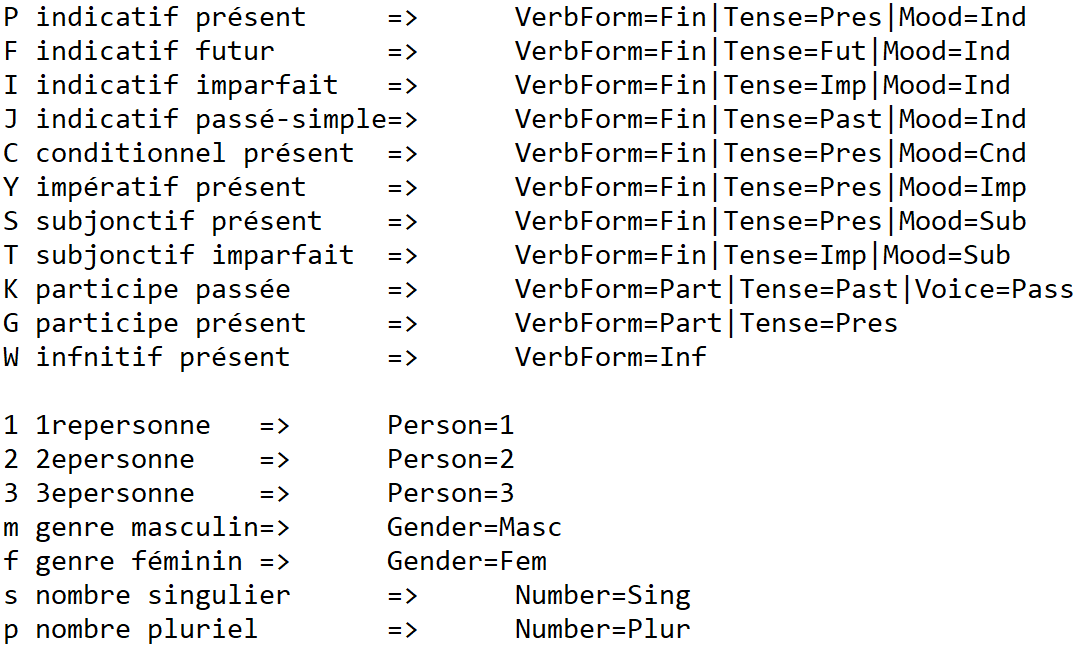
* Exemple : PSs13 = présent de l'indicatif ou du subjonctif, à la 1re ou 3e personne du singulier.

1. Comparaison et transformation les catégories et les traits en SUDfr.

* L’idée principale c’est que je veux comparer les catégories et les traits dans lefff, ud (POS tags et features) et les fichiers conll de SUDfr, puis essaie de trouver les règles qui serve à transforme directement les catégories et les traits dans lefff en forme SUD.
* Pour les catégories :

'coo' : "CCONJ", 'adv' : "ADV", 'auxEtre' : 'AUX', 'cldr' : 'PRON', 'cln' : 'PRON', 'cld' : 'PRON', 'ilimp' : 'PRON', 'prel' : 'PRON', 'clg' : 'PRON', 'clr' : 'PRON', 'cll' : 'PRON', 'pres' : 'INTJ', 'np' : 'PROPN', 'advneg' : 'ADV', 'cla' : 'PRON', 'pri' : 'PRON', 'pro' : 'PRON', 'adj' : 'ADJ', 'prep' : 'ADP', 'clar' : 'PRON', 'clneg' : 'ADV', 'det' : 'DET', 'auxAvoir' : 'AUX', 'csu' : 'SCONJ', 'nc' : 'NOUN', 'v' : 'VERB'.

* Pour les traits :



Quant aux déterminant et aux pronoms qui ont des traits comme *Definite, Degree, Poss, Reflex, PronType, etc*, ils sont difficiles d’écrire une règle commune pour la transformation, je les ajoute manuellement (les déterminants et les pronoms ne sont pas beaucoup).

Problèmes à discuter et à résoudre :

* Comme on a déjà vu dans lefff, *‘‘abaisse: PSs13 = présent de l'indicatif ou du subjonctif, à la 1re ou 3e personne du singulier’’*, il y a des mots qui présent plusieurs formes de temps et de personnes, dans ce cas-là, il faut écrire séparément ou pas ?

abaisse VERB VerbForm=Fin|Tense=Pres|Mood=Ind|VerbForm=Fin|Tense=Pres|Mood=Sub|Person=1|Number=Sing abaisser

abaisse VERB VerbForm=Fin|Tense=Pres|Mood=Ind|VerbForm=Fin|Tense=Pres|Mood=Sub|Person=3|Number=Sing abaisser

* J’ai pas encore ajouter les traits sur les nombres cardinaux et les nombres ordinaux.

(NumType = Card / NumType = Ord)

* MWEPOS (multi-word expression) : comment on va traite ce genre de traits ?
* Pour les mots ‘‘il’’ et ‘‘ça’’, est-ce qu’il faut ajouter un trait qui indique ce sont des pronoms impersonnels ?