Devoir 2 : Génération de morphologie médicale sous PC KIMMO

Le corpus de travail est une liste de termes médicaux, prétraités automatiquement pour former des couples de mots qui ne se différencient que par leur terminaison suffixale. Ils sont organisés de façon à isoler toutes les formes d'un même suffixe (= allomorphes).
On peut imaginer, par exemple, le même traitement pour les formes du préfixe IR {ir, in, im} en français :
régulier | irrégulier | | ir ; capable | incapable | | in ; mortel | immortel | | im ; étant donné le contexte droit (r, n ou m), la préfixation de ces adjectifs est soit 0, soit ir/ in/ im.

L'objectif du devoir est d'obtenir les bonnes règles de suffixation de ces termes médicaux, en vue de générer les formes suffixées à partir des bases identifiées.

Travail :

1) analyse morphologique : décrivez soigneusement les données. Repérez les « vraies » bases (qui sont parfois complexes) et les « vraies » suffixations. Pour un même suffixe, trouvez tous ses allomorphes (toutes ses variantes de réalisation), et étudiez le contexte précis de chacun. Attention, les découpages automatiques sont parfois fautifs, exemple : intra | intradermique | | dermique ; la règle morphologique de suffixation qui s'applique ici n'est pas intra | 0 | ique mais intraderme |e| ique.

2) Une fois votre description achevée, formez les règles morphologiques de suffixation. Gardez à l'esprit les questions suivantes : quelles bases acceptent quel suffixe (quelle opération suffixale) ? quel est le bon allomorphe pour chaque base (séquences épenthétiques possibles) ?

3) Formez des règles à deux niveaux permettant d'exprimer les conditions contextuelles de la réalisation de chaque allomorphe.

4) Implémentez votre travail sous PC KIMMO, de façon à générer (non à reconnaître) les formes suffixées à partir des bases identifiées.

Organisation et retour des copies :

Vous travaillerez en binôme. La liste des binômes sera relevée le jeudi 11 mai 2000. A chaque groupe sera affectée une portion de corpus, et une liste égale de suffixes à traiter. Vous trouverez sur la page web de pascal Amsili, http://talana.linguist.jussieu.fr/~amsili/, le corpus qui vous est destiné, avec votre nom.

Les devoirs doivent être rendus au plus tard le 5 juin 2000 à 10 h 00, à Pascal Amsili (dans son casier ou à son bureau en 809a.)