Hauptseminar *Using NLP to Foster Language Awareness in SLA* ISCL, Universität Tübingen
Detmar Meurers Sommersemester 2008

Exercise sheet 3
(Submit by email to dm@sfs.uni-tuebingen.de before class on Friday, June 6)

Write the little programs below, naming them nr1.py, nr2.py, etc. Save them in a folder ex3-*userid* (e.g., ex3-dm) and send me an email with a tar.bz2 archive of that directory. Please make sure that it unpacks to a directory ex3-*userid*, not the individual files.

1. Write a web form (calling a mod_python script) on which the user can enter a sentence and spell-check it. The web form should call `ispell -a` and report in user-friendly format for each word whether
   - it was in the dictionary
     - as such
     - after affix removal, based on what root
     - through compound formation
   - it was not in the dictionary
     - but there are near misses, reporting how many and which these are
     - but one could derive it by removing (illegal) affixes from a known root
     - and no near misses or derivations are available
2. Explore the nature and use of ispell dictionaries:
   - Copy a text with spelling mistakes from the Birkbeck corpus, which you can find on our aticall server under /home/dm/0643/HOLBROOKDAT.643
   - Pick three of the texts included in this file and copy them to a separate file called myfile.txt for the following tasks.

     Include a copy of the text you selected in the homework submission.
   - Run ispell over myfile.txt, noting whether it finds the errors and what suggestions it makes for fixing them.

     Briefly report and discuss in your homework what it finds.
   - Use `buildhash` to define an ispell dictionary that is restricted to the vocabulary that occurs in this article. Using this dictionary, run ispell over myfile.txt noting whether it finds the errors and what suggestions it makes for fixing them.

     Include the source of the dictionary you created. Report and discuss what ispell now finds as errors and makes as suggestions, comparing it to the unmodified ispell result.