

Big Homework 2

Human Language for Artificial Intelligence: 11-324, 624, 724

Due Friday, October 27, 2023, 11:59pm

1 About the assignment

Big Homework 2 requires you to create an FST for Somali morphophonology. The overall directions for the assignment can be found on pages 55-57 of the Levin and Mortensen HL4AI textbook draft.

2 What to do and what to hand in

1. Begin by determining the underlying representations for each of the roots and each of the suffixes. To that end, submit a file containing:
 - a list of underlying representations you posit for each root
 - a list of underlying representations you posit for each suffix
 - any commentary on how you determined the underlying representations and how you ordered the rules
2. Now that you have the underlying representations for both the roots and the suffixes, create a list of posited underlying forms for each word. You will use this list in step 4.
3. Determine the rules producing the surface representations.
 - Submit a foma script (*somali.xsft*) defining an FST that transduces between Somali underlying representations and surface representations.
4. Run your *somali.xfst* script. Make sure to add your text file containing the underlying forms you posit for each word.
5. Output the surface forms using the command *print lower-words*.
6. Verify that all the transduced surface representations are correct against the original provided data.
7. Zip all files

- file containing underlying representations posited for roots and suffixes
- .xsft script
- file containing your posited underlying forms transduced into their surface representations

into a file named family-name_given-name_bhw2.zip – e.g. *gish_caroline_bhw2.zip*.

3 The Data

The Somali data for which you will be writing rules can be found on pages 55 and 56 of the Levin and Mortensen HL4AI textbook draft.

The data contain singular, singular definite, and plural nouns as well as third person singular masculine and third person singular feminine verbs. Though there are seemingly three different sets of data in the book and 2 separate downloadable data files, you do not need to create separate files of rules or lists of underlying forms.

4 Hints

- Total of 45 posited underlying root representations
- Total of 6 posited underlying suffix representations
- Total of 135 underlying forms