# ANLP Self-Assessment Exercises for Week 2 (v1.5)

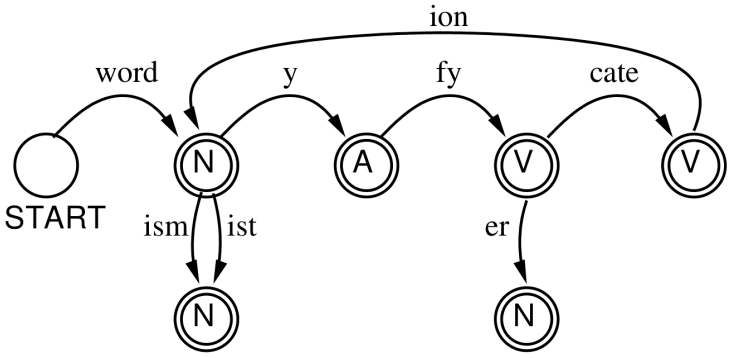
Adam Lopez, Sharon Goldwater
(School of Informatics, University of Edinburgh)

## Goals

To give you practice working with finite state machines and thinking about these as an example of a model of language: what does a particular FSM predict about the language and why is the model designed the way it is?

## Exercise 4

In lecture 3, there was an example of English derivational morphology based on the word *word*. The FSM below implements a fragment of English morphology that looks like this and generates words like *wordy*, *wordification*, etc. (assuming that spelling changes are fixed up by another FSA that applies afterward).




a) The example only includes a single stem, *word*, on the first transition arc. List three other stems that could go there. What kinds of words

*can't* go there?

b) Consider the transitions labeled *er*, *ism*, *ist*. All of these end up in states labelled *N* (noun). What would happen if we removed the bottom two noun states and made these transitions end up in the same state where the *word* transition ends? Give some examples of words that are generated. Do these seem like possible words of English to you? (You might have different judgments than other people!)

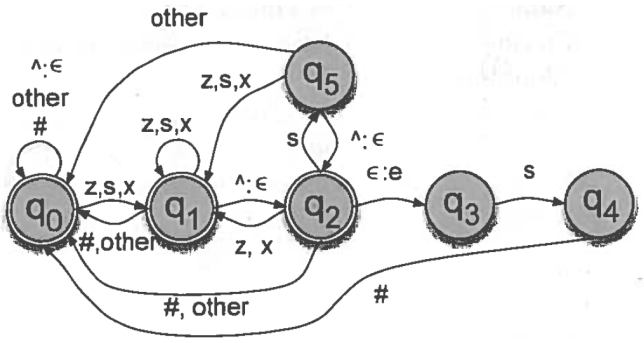
**Solutions**

a) A few other stems that could go there: *Thatcher*, *smell*, *fuzz*. Any word that isn't a noun cannot go there. I haven't been able to think of any noun stems that can't, but maybe someone in the class can! As for nouns in general, it may depend on your judgment as to whether nouns with existing endings can go there or not (see also next question): for example my own judgement about *wordism+y* is a valid word is a bit fuzzy.

b) We'd get words like *wordismification*, *wordifierify*, *wordisty*, etc. I think these are legal, but pretty weird. Your mileage may vary.

## Exercise 5

The transducer from J&M Fig 3.17 is reproduced below. ('other' = none of {z,s,x,ˆ,#,$\epsilon$}).




a) What sequence of states would we go through to create the correct plural form for `axleˆs#`?

b) What about for `lassˆs#`?

c) Can you think of any words that cause the transducer to go from state $q_2$ to $q_5$ and then continue on to an accepting (end) state? If not, can you at least say what properties would such a word need to have?

**Solutions**

a) `axle^s#` (showing also the consumed symbol to reach each state): Start in $q_0$, then move to: $q_0$ (`a`), $q_1$ (`x`), $q_0$(`l`), $q_0$ (`e`), $q_0$ (`^`), $q_1$ (`s`), $q_0$ (`#`)

b) `lass^s#`: Start in $q_0$, then move to: $q_0$ (`l`), $q_0$ (`a`), $q_1$(`s`), $q_1$ (`s`), $q_2$ (`^`), $q_3$ ($\varepsilon$), $q_4$ (`s` , $q_0$(`#`) Notice that at some point(s) there is more than one option where to move (nondeterminism). To create the correct plural, you need to choose the right option(s). In a nondeterministic FST, if *any* path leads to an accepting state, then the string is accepted. To implement this kind of nondeterministic FST in practice, we would need to use an algorithm to convert it to a deterministic FST.

c) The word would need to have a morpheme boundary preceded by one or more z's, x's, or s's, and also followed by an s and some additional characters. So the 's' after the morpheme boundary in this word couldn't be a plural 's' but the start of some other morpheme.

This made me think about compound words. The only one I've thought of so far is a bit contrived: `fox^sit#` (a verb meaning to watch someone's pet fox while they're out, similar to *babysit* or *dogsit*).

A student also pointed out that prefixed words will work too, and are less weird. For example `dis^similar`, `dis^satisfy`.

Of course, these examples are not plural forms, so it's not clear whether we'd ever send these words through this transducer as part of a larger system. This transducer clearly is not sufficient to deal with all possible spelling changes in English (e.g., sending it `skip^ing` doesn't output the right answer), but if you imagine that it's part of a larger construction that deals with those other aspects, then non-plural words would need to be considered as possible inputs.