Linguistics 219
Phonological Theory III

Winter 2018
B. Hayes

Homework #2: Biased modeling of generality effects in maxent

Due Monday April 23 in class

1. Rationale

This problem addresses an issue confronted in non-OT terms by Albright and Hayes (2003; *Cognition*; linked from course web site). It has to do with the effect of *degree of attestation* of morphological patterns on the degree to which speakers take them seriously when they wug test.[1] The Albright/Hayes study was a wug test on English past tense forms, together with a comparison of models on how to predict the results. We were particularly interested in "islands of reliability", phonological configurations in which certain past tense changes are especially common, perhaps productive.

Here are some crucial data patterns.

- In English, *every single verb that ends in Clɪŋ* has a past tense of form Clʌŋ (*fling* ~ *flung*, *cling* ~ *clung*, see spreadsheet for further examples. (Things are almost as unanimous if we include the other liquid [r]). Yet when the experimental participants were wug tested on *spling*, they preferred *splinged* to *splung* by a ratio of about 61% to 39%.[2] In other words, the experiment observed not frequency matching, but instead the effect of some sort of bias — for the most common pattern? for the pattern with no stem alternation? Hmm …
- About half of the verbs in […ip] takes past tenses in […ɛpt] (*keep* ~ *kept*, *leap* ~ *leapt*, etc., but *seep* ~ *seeped* etc.). Yet the Albright/Hayes experimental subjects preferred *fleeped* to *flept* by .77 to .24, not 50/50. Note that the toleration of irregulars is lower, compared to the Clʌŋ verbs, perhaps because in this case there exist exceptions to the special past tense pattern.
- Lastly, consider *bring* ~ *brought*. This case is essentially unique[3] and the general finding (Albright/Hayes, others) is that such single-example cases lead to almost *no* analogous responses. For instance, Albright and Hayes got almost zero cases of *zed* [zɛd] for *zay* [zeɪ], which would follow the one-example model of *say* [seɪ] ~ *said* [sɛd].

Suppose we set up a constraint-based morphology that had **mapping constraints** that favored particular string mappings in forming outputs, like these:

---

[1] By which I include everyday-life wug testing, common for kiddies.

[2] I hereby discard the *splang* responses for simplicity.

[3] There is *catch* ~ *caught*, forming what is sometimes called an "output-oriented generalization" (see work of Bybee, Kapatsinski).

FORM PAST TENSE ON C{l,r}ɪŋ VERBS BY CHANGING [ɪ] TO [ʌ]
FORM PAST TENSE BY SUFFIXING /-d/ AND DOING PHONOLOGY.[4]

These would form conflicting constraints, and adopting standard learning principles *both classical OT and maxent would fall flat on their faces*, favoring whatever prevails in cases of conflict. Thus *spling* could *only* be made past tense as *splung*, and native speakers would consider *splinged* grossly ungrammatical. This prediction is clearly wrong.

2. Assignment: background

Download the spreadsheet that accompanies this homework. It includes a simplified overview of English past tenses, including the cases described above plus a large number of regulars. Some of the regulars conflict with irregular patterns, like *heap*, whose candidate *heaped* violates the mapping constraint for [Cip] ~ [Cɛpt]. Others, like *fuss*, violate no other mapping constraints, and they are listed as a big, trivial, no-rival-candidate bundle.

At the end of the spreadsheet are wug forms from Albright and Hayes, with observed frequencies taken by averaging their wug tests. The frequencies are *not* included in the harmony computations of the main spreadsheet; rather, they serve a test for whatever *biased* learning system you may try to install. Of these, *splung* tests best and *flept* not as high.

I include a complete fictional wug form, BRING. This is meant to represent how people would respond to *bring* if they didn't already know that its past tense is *brought*. This is not entirely crazy, since little kids learning English almost invariably go through a phase of giving its past tense as *brung* or *bringed*.[5]

3. Assignment: instructions

First, reassure yourself that if you weight the constraints without bias, they will match the observed data correctly, and the wug test data very poorly. Report this result.

Then, you are on your own. You must invent a biased learning procedure that takes in the raw empirical forms at the top of the spreadsheet, arrives at a grammar, and then projects something like the wug-test frequencies.

My own solution follows the course material: I set up mu's and sigma's, letting them vary as a function of *the number of forms in which a particular mapping is attested*. This let me penalize the irregular mappings relative to the regular mapping, and also penalize the unique *bring* → *brought* mapping relative to the two others. Note that this solution is based solely on numbers, and makes no use of the fact that the regular mapping has no stem alternation.

---

[4] For some possibilities of doing morphology, see Alex MacBride's UCLA dissertation (on our web site) and work he cites.

[5] A great source for this is Marcus, Pinker et al. (1992) *Overregularization in language acquisition*, Society for Research in Child Development.

But you can invent anything you like. I believe that formulating a good model of Generality Bias is one of the great unsolved problems of linguistics, so see what you can come up with!

Feel free to contact me with questions.