Massachvsetts Institvte of Technology
6.S077, Fall, 2017
Reading and response #1
Learning Language by Statistical Methods

Released: Wednesday, September 20 **Discussion in class**: October 4/6

Goal: Learn how to write a succinct scientific review; learn about word segmentation in language

Reading:

Saffran, Aslin, and Newport, Statistical learning by 8-month-old infants, 1996. *Science*, 275:1926-1928, available via the 6.034 web page as:
http://web.mit.edu/6.034/www/6.s077/saffran96.pdf

Assignment

You are a reviewer for the prestigious journal *Science.* You've just received a startling submission, from the Newport Lab at Rochester University. It's startling because it claims that very young infants might not need as much 'innate knowledge' as has sometimes been asserted. Instead, infants might be able to figure out where word boundaries are from statistical regularities alone. It's your job to write a pointed, 1 or 2 page review that says either: (1) "Accept as is, no changes needed"; (2) "Accept, but with the following minor revisions;" (3) "Accept with the following major revisions;" or (4) "Reject". **Please place your evaluation at the very top of your review.**

Naturally, you must also justify your evaluation. Also, you must also be extremely precise and pithy; the *Science* editor doesn't have a lot of time to wade through a review that simply re-states what the submission is about – time and space are precious. So, in particular you decide to focus on a few key questions and on the central experiment that they did: you pay close attention to footnotes 19, and 20 that describe the training and test stimuli that were set up in order to determine whether infants were, in fact, *learning* and then *using* transition probability information to determine 'word' boundaries after only 2 minutes exposure to an unbroken stream of test stimuli. *Please state why or why not you think they established their hypothesis*.

In addition to any justification for your review, please answer these following questions in your submission, since we'll call on folks to reply to them during the classroom part of the assignment. (It is helpful then to write these answers down separately from your review, but you can weave the results back into your review to hand in; we just want you to be able to refer to the answers quickly when you are called on.)

1. In a single short sentence, please state clearly what *hypothesis* Saffran *et al*. were trying to establish.

2. What knowledge do Saffran *et al.* assume 8-month babies already know before they are exposed to the experimental stimulus? What knowledge do Saffran *et a*l. say is acquired by the infants after exposure to the experimental stimulus? (We want you to be able to briefly describe the initial and the final state of knowledge the babies had, relative to this

particular experiment–the key to describing any learning system.)

3. The experiment was designed so that the training data presented to the infants consisted of four different nonsense 'words' strung together. Each fake 'word' was made up of 3 possible Consonant-Vowel (CV) combination 'syllables,' so each word consisted of the following Consonant-Vowel pairs: CVCVCV, e.g., '`pabiku`'. Then all the trisyllabic 'words' were run together to create a sound stream without any breaks to form a sequence of training data. The training data was constructed in a particular way to avoid the possible influence of 'other factors' that could make it difficult to ascertain the main effect that Saffran *et al.* wanted to investigate. In a sentence or two, please describe a few of these 'other factors.'

4. What purpose did their "condition B" serve, as described in footnote 20?

5. In our segmentation analysis program below, we report both *precision* and *recall* values. We define *precision* as: (true positives)/(true positives + false positives) and *recall* as (true positives)/(true positives + false negatives).

   Please answer the following two questions:

   5.1 Why do we calculate both *precision* and *recall* than simply calculating the total number of correct answers that the trained system makes, i.e., the *accuracy*, defines as (true positives + true negatives/(true positives + true negatives + false positives + false negatives)?
   5.2 Suppose the correct target segmentation of *bigbadwolf* is as three separate words *big*, *bad*, and *wolf*. Now suppose the child segments *bigbadwolf* as *bigbad* and *wolf*. What is the child's precision and recall in this case?
   5.3 Explain how it is possible to push *precision* to a high value, close to 100%, by the careful selection of training and test examples, and why that would typically lower *recall*. Then explain how it is possible to do the same thing with *recall*, pushing it to nearly 100% while at the same time lowering *precision*. (As a result, researchers typically report both precision and recall values or combine them via some weighting scheme.)

6. How well does the Saffran *et al.* "transitional probability" method work on actual mothers' speech to children? In class, we indicated that it did not work very well, but it is more fun to try this out for yourself using a computer program and real data. We have provided a zip file under the **`6.S077/Assignments/`** section on the 6.034 wiki page as **`seg.zip`**. Unzip this file, and you'll have two files: (1) a python program **`spacestp.py`** and (2) a data file **`mother.speech.txt`**. This last file is a transcription into phonetic form (along with stress markings) of a real corpus of mothers' speech to their kids, using the CMU pronouncing dictionary located here:

   `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`

   The text has been marked up with each end of a word labeled as W; the end-of-utterance

(end of sentence), by U; each syllable by S, each individual sound or phoneme by P, and stress by a number, with 1 = primary stress. For example, the first four sentences the mother says to their child Adam are: "Big drum. Big drum? Horse. Who is that?" This is transcribed into an unbroken stream as follows:

bPih1PgPSWdPrPah1PmPSWUhhPao1PrPsPSWUhhPuw1PSWih1PzPSWdhPae1PtPSWU

For example, the first utterance "big drum" is **bPih1PgPSWdPrPah1PmPSWU**, which comes out of the CMU system as: **b ih1 g dr ah1 m**. (Try it yourself on the CMU website.). Then the file has added the P, W, and U markers as described. The **spacestp.py** program uses both the Saffran *et al.* transitional probability method and other methods to discover where the word boundaries in the corpus are (of course, it first strips off the "W" marks, because to use them would be cheating). If you run the program:

```
python spacestp.py
```

then it will process the file **mother.speech.txt**. Report back its success in using both transitional probabilities and stress in finding word boundaries. It success is reported in terms of precision and recall.

Please run this program and turn in the values you get as part of your report. Then, in a sentence or two, please explain *why* it is that stress seems to do so much better at finding word boundaries than transitional probabilities–at least in terms of precision–in this actual sample of mothers' speech to children.