This handout includes space for every question that requires a written response. Please feel free to use it to handwrite your solutions (legibly, please). If you choose to typeset your solutions, the `README.md` for this assignment includes instructions to regenerate this handout with your typeset LaTeX solutions.

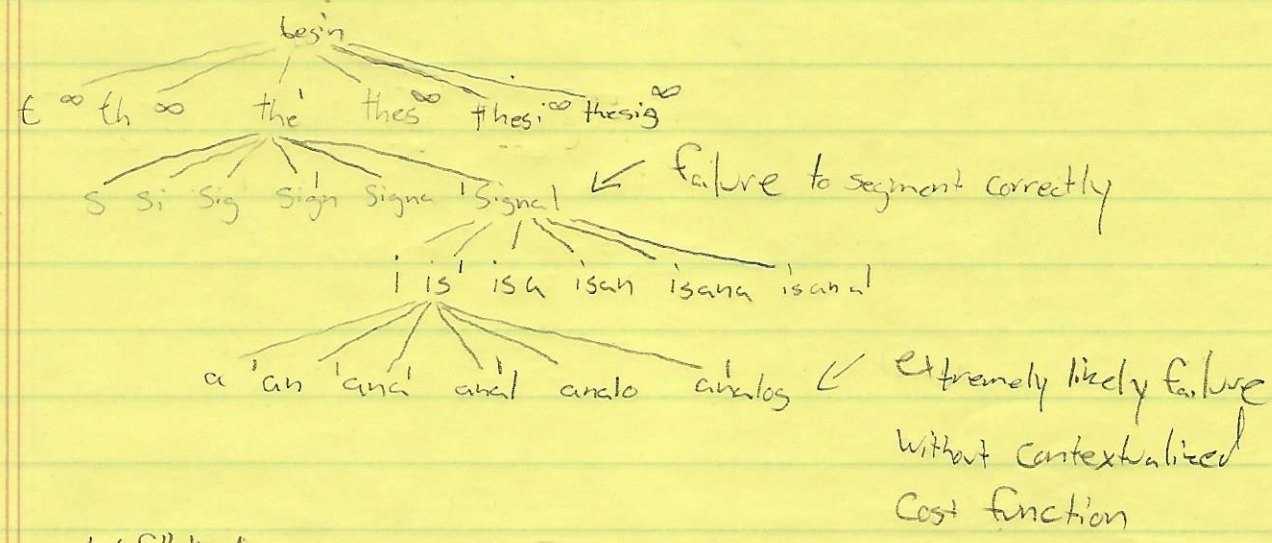
1.a

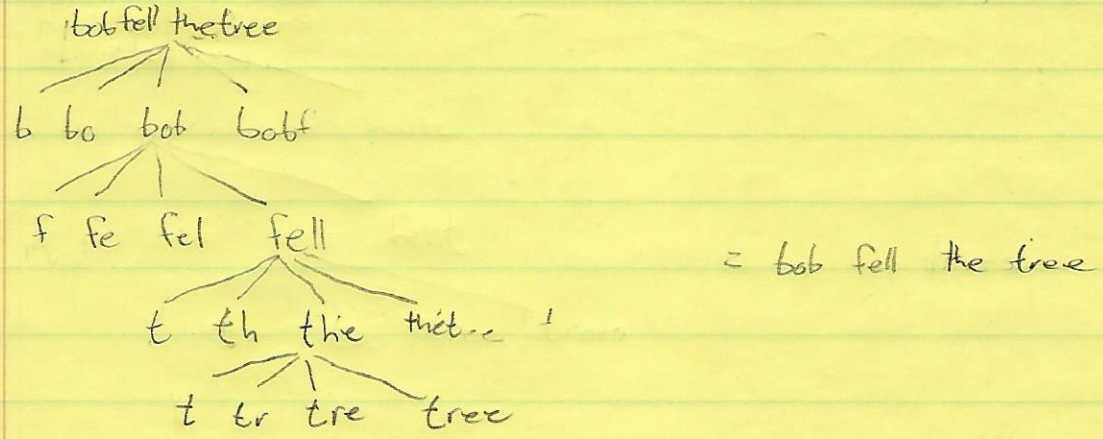
See image next page

1(a) Word Segmentation

Unigram model with a binary cost function $C(w_i) = \begin{cases} w_i \text{ in dictionary} = 0 \\ w_i \text{ not in dictionary} = \infty \end{cases}$

Query = "thesignalisanalog"





2.a

A good example sentence where a greedy vowel insertion implementation would be challenging is in the sentence "the mood is bright" = "th md s brght".

A reasonable set of possibleFills here are [(the),(mad,mood,mud,meed,mood),(as,is,us),(bright, brought)]

The real challenge here is selecting mood with a gready implementation. The bigram with "the" would likely be somewhat equally weighted across the 5 options listed.

3.a

State space
(Location in the string we are presently looking at, Prior word selected)

Actions
Move from the previously selected word to lowest cost possible fill of the shortest segment of the string assuming that it is the only option

Costs
Bigram correlation score

End test
Reaching the end of the string

3.c

State space
Location in the string we are presently looking at

Actions
Move from the previously selected word to lowest cost possible fill of the shortest segment of the string assuming that it is the only option

Cost
Bigram correlation score plus the difference between the heuristic of the end state and initial state

End state
Reached the end of the string

h(s)
A singular weight vector that matches the dimensionality of the embedded version of the current state

For the Heuristic to work h(s) must not exceed the total cost of the inserted and segmented string to ensure we don't go negative for UCS. To achieve this the heuristic of the desired end state needs to be 0. This is achieved during training of the weight vector.

4.a

4.b

4.c