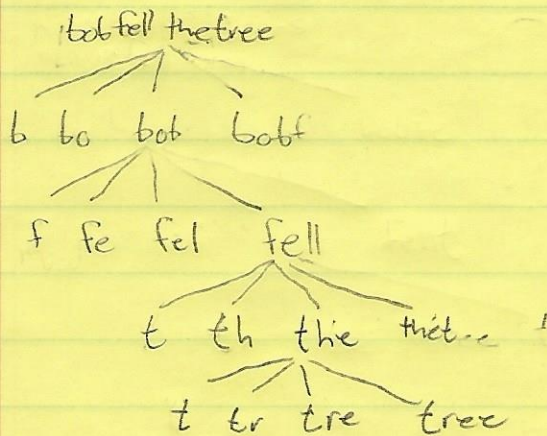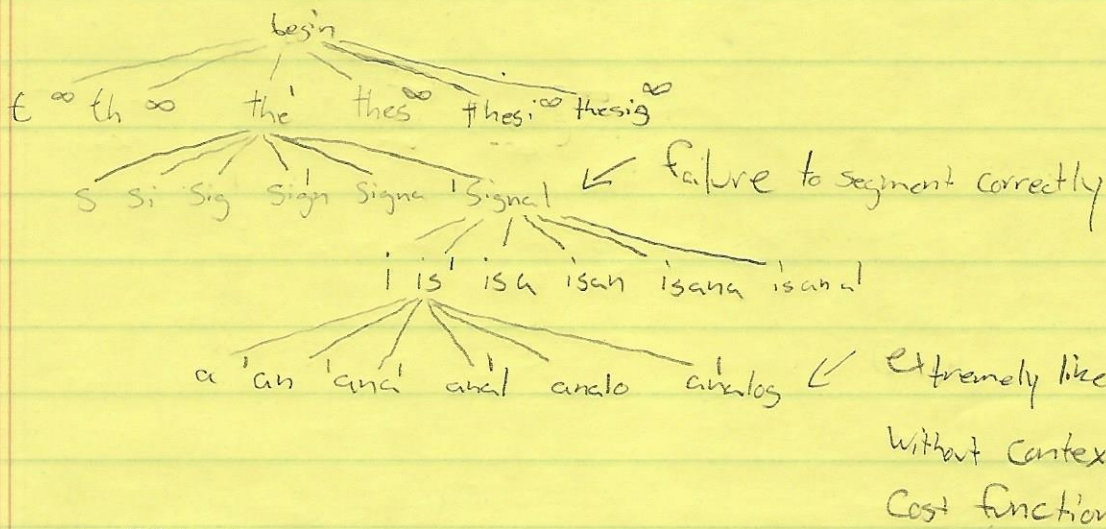# 1(a)   Word Segmentation

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"

begin

t ∞ th ∞ the'   thes ∞ thesi ∞ thesig ∞

s   si   sig   sign   signa   'signal   ← failure to segment correctly

'i   is'   isa   isan   isana   isanal

a   'an   'ana'   anal   analo   analog   ← extremely likely failure without contextualized cost function

bob fell thetree

b   bo   bob   bobf

f   fe   fel   fell

t   th   thie   thet...

t   tr   tre   tree

= bob fell the tree

2(a)   Vowel Insertion

the mood is bright

th md s brght

(the) (mad, med, mud, meed, mead, mood)

           mood ←(as, is, us)

                   is   (bright, braght)

           the mood is bright