

1.a

Input string, "tryit sometime"

Unigram cost function:

Word	Cost
try	1
it	1
some	3
sometime	4
time	3
other	10

As to the greedy algorithm we will get: try it sometime. However the optimal output would be:

'try it sometime' with less cost.

"Was the greedy search not always provide optimal solution.

2. a

Let's say we have the following string input: "you ready?" with bigram cost function and possible fills:

" → are, our

y → you, ya

rdy → ready

cost(-begin-, are) = 4

cost(-begin-, our) = 3

cost(are, you) = 1

cost(our, you) = 6

cost(are, ya) = 10

cost(our, ya) = 10

cost(you, ready?) = 2

cost(ya, ready) = 10

The greedy algorithm would return "our you ready?" with cost 11 while the optimal

case is, are you ready?" with
cost 7.

3. a.

- State in this problem is a tuple of (`remainText`, `previousWord`)
- The start state would be queryWord,
`SENTENCE-BEGIN`.
- The end state is when `remainText == 0`.
(is empty)
- Actions:
 - applying segmentation to `remainText` (generate all possible words)
 - For each word generated, insert vowels. If such word exists, compute bigramCost (`previousWord`, `vowelInsWord`)

· New state : (remainTextSegment,
vowelInWord).

3. C.

Define $f_U(w) = \min_b b(w^*, w)$ \rightarrow
the cost of unigram model is
smaller or equal to any of the
bigram models.

Define a relaxed search problem
 P_{rel} to be : state = index
cost is using unigram model.
Thus from definition of the unigram
model the cost will be smaller or
equal to the original problem.
It means it is a relaxed problem,
so estimation of the heuristic is
consistent.

4. a

Only the first example was reconstructed correctly because it's bigram occurred in given training data

4.b.

This algorithm can reconstruct only texts written in classic English. It doesn't work with text which includes modern phrases, expressions and slang language.

4. c

To improve this algorithm we need to increase training data or provide another training set which will includes word, phrases

of modern use. That will help to decrease current limitations and bigram cost so it will allow us to use this algorithm to trigger (wider) variety of texts.