1 – a:

This greedy algorithm may result in suboptimal solutions because for example:

“andgodsaidbehold”

Cost:

and	1
go	2
god	10
said	4
be	5
hold	6
behold	10

Using the greedy algorithm, we’ll get:

“and god said be hold”, total cost = 1+2+10+4+5+6 = 28

But the optimal solution is:

“and god said behold”, total cost = 1+2+10+4+7 = 27

And our assumption of the costs make sense, because “be” is very common, while “behold” is fairly uncommon. However, it could still happen that the cost of “behold” is larger than the costs of “be” and “hold” combined. In this case, this greedy algorithm is suboptimal.

2-a:

This algorithm is suboptimal as shown in the following example:

“bm m p”

Possible fills:

fg	fugue fog
bm	beam boom
m	me am
p	ip up

Costs (bigram):

beam me	3
beam am	1
boom me	3
boom am	10

me ip	5
me up	2
am ip	5
am up	5

If we use the greedy algorithm, we will get:

“beam am ip” or “beam am up” => score = 1+5 = 6

But apparently the best solution is:

“beam me up” => score = 3+2 =5

The algorithm could generate suboptimal results if there exist very low cost bigrams ensuing high-cost initial bigrams.

3-a:

State: (Last whole word, Index of the last space insertion). Last whole word is defined as the last word that had its vowels filled and space inserted.

Action: Last whole word

Costs: The bingramCost() of (Last whole word, current word from possibleFills())

Initial state: (-BEGIN-,0)

End test: Index of the last space insertion == len(query)

This problem has the same start and end states as 2-b. Some code can be reused.

3-c:

We want to get rid of w’ in our cost function, therefore:

$\mathbf{u_b}$ = min_{w’} b(w’,w). We know that by this definition, $\mathbf{u_b}$ <= b(w’,w) at all times.

Relaxed problem:

State = i, the index of the last space insertion (i.e. i+1 is the starting point of the current word w)

Cost(i) = $\mathbf{u_b}$(w)

Proof of consistency:

The modified cost, $\mathbf{u_b}$, is always >= 0, since b(w’,w) >= 0.

Intuitively, the heuristic of the end state is 0 since there’s no future edges thus no future cost.