1. The Multiple-stack Algorithm as A* Search: Show that the multiple-stack algorithm of Section 6.6.1 is an instance of the generic A* algorithm.

Specifically, the A* search in Section 6.3 is concerned with finding the best hypothesis \( \hat{w}^n \) among all hypotheses \( w^n \) of a given length \( n \). The parameter \( k \) is associated with the length of a partial hypothesis \( w^k \), and a goodness function \( F(w^k) = g(w^k) + d(w^k) \) is defined for each partial hypothesis. Starting with the empty hypothesis in a stack, the current best partial hypothesis \( \hat{w}^k \) is extended at each iteration to \( \hat{w}^{k+1} = \hat{w}^k \| \hat{w}_{k+1} \), until the best partial hypothesis attains a length \( n \).

However, given the acoustics \( a^m = a_1, \ldots, a_m \), the goal in Section 6.6.1 is to find \( \hat{w}^n = \arg\max_w P(a^m, w) = \arg\max_w \log P(a^m, w) \), (1) among all word sequences \( w \) regardless of their length, and \( n \) is not given. This exercise is about extending the notation of Section 6.3 to solve the problem (1) stated above.

To this end, define a partial hypothesis to be a pair \( \langle l, w^k \rangle \), where \( 0 \leq l \leq m \), and let

\[
g(\langle l, w^k \rangle) = \max_{l_1, \ldots, l_{k-1}} \sum_{i=1}^k \log P(a_{l_{i-1}+1}, w_i),
\]

where \( 0 = l_0 < l_1 < \ldots < l_k = l \).

Next, define the “length” of a partial hypothesis \( \langle l, w^k \rangle \) to be \( l \), not \( k \), so that the length of a complete hypothesis is \( m \). Extending an \( l \) length hypothesis to a \( l + 1 \) length hypothesis therefore does not necessarily entail extending \( w^k \) to \( w^{k+1} \): we could extend \( \langle l, w^k \rangle \rightarrow \langle l + 1, w^k \rangle \) just as easily as \( \langle l, w^k \rangle \rightarrow \langle l + 1, w^{k+1} \| w_{k+1} \rangle \).

Finally, to show that the multiple-stack algorithm is an A* algorithm, let

\[
d(\langle l, w^k \rangle) = \max_z \max_{l_1, \ldots, l_{|z|-1}} \sum_{i=1}^{|z|} \log P(a_{l_{i-1}+1}, z_i), \quad l = l_0 < l_1 < \ldots < l_{|z|} = m,
\]

where \( |z| \) is the length of the word sequence \( z \). Note that \( d(\langle l, w^k \rangle) \) only depends on \( l \).
(a) Show that if an A* search is conducted using a single stack, with

\[ F(\langle l, w_k \rangle) = g(\langle l, w_k \rangle) + d(\langle l, w_k \rangle) \]

as the goodness of a partial hypothesis \( \langle l, w_k \rangle \), and is stopped the first time an \( m \) length hypothesis \( \langle m, \hat{w}_1^n \rangle \) percolates to the top, then \( \hat{w}_1^n \) is indeed the most likely word sequence.

(b) Derive a relationship between \( d(\langle l, w_k \rangle) \) and \( g^*(l) \) as defined in Eqn (11) on p100.

(c) Show that sorting the stack entries during this A* search according to

\[ F^*(\langle l, w_k \rangle) = g(\langle l, w_k \rangle) - g^*(l), \]

will also lead to the discovery of the same \( \hat{w}_1^n \); i.e. \( \langle m, \hat{w}_1^n \rangle \) will be the first \( m \) length hypothesis to percolate to the top of the (single) stack.

Argue based on these results that when the multiple-stack algorithm on p101 stops, the top entry \( \hat{w}_1^n \) in the \( m \)-th stack is the word sequence that maximizes \( g(\langle m, \cdot \rangle) \) in (2).

2. **N-best Paths Using the Multiple-stack Algorithm**: Modify the multiple-stack algorithm of Section 6.6 to obtain the \( N \) best hypotheses instead of only the best. In particular, assume that the conditions (a), (b) and (c) of Section 6.6.1 on p100 are satisfied.

(a) Minimally modify the algorithm of Section 6.6.1 to obtain the two best paths.

(b) Check if the extension of Section 6.6.2 will hold for your new algorithm.

(c) Generalize your modification for any \( N \geq 2 \) best paths.

Discuss whether the extension to the actual multiple-stack algorithm of Section 6.6.3, when the assumptions (a), (b) and (c) do not hold, will be possible for your new algorithm. What are the pitfalls for \( N \geq 2 \) that were not there for \( N = 1 \)?