NLP – EX1

1. The probability of never generating stop is:

Which is a Bigram model is equivalent to:

Since for each word :

And the transition probabilities sum to 1:

Therefore:

And

1. 1. Unigram model:

Where

and

If the word ‘where’ appeared more times in the corpus than the word ‘were’

The sentence with the highest probability will be:

He went were there where more opportunities.

(and vice versa)

a mixture of both options won’t receive a high probability in this model, and thus it will never give the correct sentence the highest probability (unless both appear the exact amount of times, under which the behavior in undefined).

* 1. Bigram model:

This model can give the correct sentence the highest probability, because the assumption that ‘there were’ appears more in corpus than ‘there where’

And |went where| is likely higher than |went were|, thus their probabilities will be larger.

We do risk the possibility that a given two-word sequence in the sentence will not appear in the corpus, and in such a scenario the sentence probability will be 0.

1. 1. We need to show that:
2. 1. Trigram model:

The made assumption is that each word appearance is only dependet on the two previous words, and in independent from the rest.

* 1. In case of emergency, the child hastily fastens the safety belt.

זה המרצה הכי מעניין בתואר!

* 1. In case of emergency, children should hastily fastens the safety belts.

הכלבה היפה שראינו נובח.

We would need a 4-Gram model for capturing subject-verb agreement in both sentences.

1. Pairs:

ראיתי איש הלך.

Triplets:

ראיתי איש הולך ראיתי.

4-Tuples:

ראיתי איש הולך למכולת ראיתי.