

**Init:**

Go through all available training data (labeled and unlabeled) and obtain frequency counts for every n-gram ( $n$  in  $\{1, 2, 3, 4\}$ ) that occurs.

$sortedNGrams \leftarrow$  Sort n-grams by frequency in descending order.

**Loop**

until stopping criterion (see Section 3.3) is met

1.  $trigger \leftarrow$  Go down  $sortedNGrams$  list and find the first n-gram that isn't covered in the so far labeled training data.

2.  $selectedSentence \leftarrow$  Find a sentence that contains  $trigger$ .

3. Remove  $selectedSentence$  from unlabeled data and add it to labeled training data.

**End Loop**