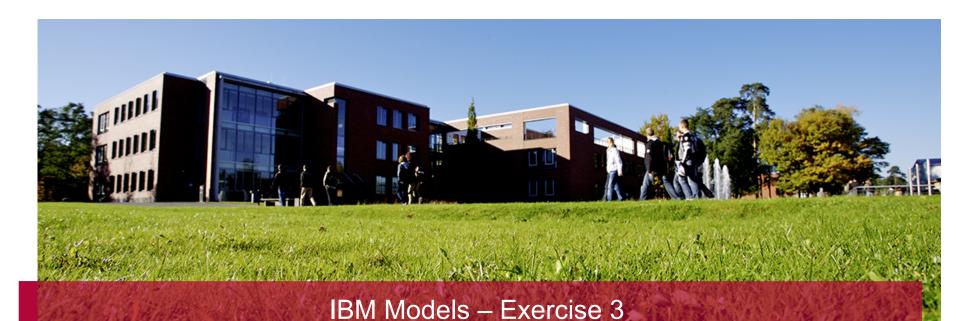
Machine Translation WiSe 2015/2016

Dr. Mariana Neves



IT Systems Engineering | Universität Potsdam

November 9th, 2015





Goal

Implement word-based IBM Models



Exercises

- Model 1 (mandatory)
 - Deadline together with exercise 2 (Nov, 22nd)



Algorithm for Model 1

```
Input: set of sentence pairs (e, f)
                                                    14:
                                                                // collect counts
                                                                for all words e in e do
Output: translation prob. t(e|f)
                                                    15:
 1: initialize t(e|f) uniformly
                                                    16:
                                                                   for all words f in f do
                                                                      count(e|f) += \frac{t(e|f)}{s-total(e)}
 2: while not converged do
                                                    17:
 3:
        // initialize
                                                                      total(f) += \frac{t(e|f)}{s-total(e)}
                                                    18:
      count(e|f) = 0 for all e, f
                                                    19:
                                                                   end for
 5:
        total(f) = 0 for all f
                                                    20:
                                                                end for
 6:
        for all sentence pairs (e,f) do
                                                    21:
                                                            end for
 7:
           // compute normalization
                                                    22:
                                                            // estimate probabilities
 8:
           for all words e in e do
                                                            for all foreign words f do
                                                    23:
               s-total(e) = 0
 9:
                                                    24:
                                                                for all English words e do
               for all words f in f do
10:
                                                                   t(e|f) = \frac{\operatorname{count}(e|f)}{\operatorname{total}(f)}
                                                    25:
11:
                  s-total(e) += t(e|f)
                                                    26:
                                                                end for
12:
               end for
                                                    27:
                                                            end for
13:
           end for
                                                    28: end while
```



Exercises

- Model 2 (optional)
 - 0.5 point in the exam's grade (20 points)
 - Deadline together with exercise 2 (Nov, 22nd)
- Model 3 (optional)
 - 1.0 point in the exam's grade (20 points)
 - No deadline, until the last week of the course (Jan, 31st), but should also include presentation.
- Algorithms for Models 2 and 3 are presented in the book



Submission

- Deadline (Model 1, mandatory)
 - Sunday, Nov 22rd, 23:59
- Hand-in
 - SQL file with queries
 - Stored procedures in the team's schema
- Presentation
 - Monday, Nov 23rd (together with exercise 2)
- Not graded, but all teams need to present at some point