

# Fortgeschrittene Methoden der statistische maschinelle Übersetzung

## Übung 2

Abgabe (pdf, Zip) an [kaeshammer@phil.uni-duesseldorf.de](mailto:kaeshammer@phil.uni-duesseldorf.de)  
bis einschließlich Donnerstag, 20.11.2014

### Aufgabe 1

Lösen Sie die folgende Aufgabe auf Papier oder durch Implementieren des Algorithmus in Python 2.x (oder beides ;)).

Gegeben ist ein Korpus mit drei Satzpaaren  $(\vec{e}, \vec{f})$ :

1. (the dog barked , bellte der Hund)
2. (dog , Hund)
3. (the table , der Tisch)

$V_E = \{\text{barked, dog, table, the}\}; V_F = \{\text{bellte, der, Hund, Tisch, NULL}\}$

(a) Berechnen Sie die lexikalischen Übersetzungswahrscheinlichkeiten  $t(e|f)$  (für alle  $e \in V_E, f \in V_F$ ) für IBM-Model 1 nach drei Runden EM-Algorithmus. Vergessen Sie die Null-Alignierungen nicht!

(b) Geben Sie für jeden Satz im Korpus und außerdem für das Satzpaar (the table barked, der Tisch bellte) die wahrscheinlichste Alinierung auf Basis dieser Übersetzungswahrscheinlichkeiten  $t_3(e|f)$  an.

Falls Sie sich für eine Implementierung entscheiden, dürfen Sie auch gerne mit mehr als drei Iterationen bzw. mit anderen Daten herumspielen.

### Allgemeine Kommentare zu einer Implementierung

- Fügen Sie bitte eine Readme-Datei hinzu, die spezifiziert, wie der Code aufzurufen ist.
- Stellen Sie sicher, dass ein Außenstehender Ihren Code versteht, zum Beispiel durch Kommentare.

- Die Python Standard Library (<https://docs.python.org/2/library/index.html>) dürfen Sie gerne verwenden, zum Beispiel für Mengen, Dictionaries (Assoziative Arrays) usw.
- Code von Programmen, die beim Ausführen Fehler erzeugen, wird nicht akzeptiert.

## Aufgabe 2

$f = \text{das}$		$f = \text{Haus}$		$f = \text{ist}$		$f = \text{klein}$	
$e$	$t(e f)$	$e$	$t(e f)$	$e$	$t(e f)$	$e$	$t(e f)$
the	0.7	house	0.8	is	0.8	small	0.4
that	0.15	building	0.16	's	0.16	little	0.4
which	0.075	home	0.02	exists	0.02	short	0.1
who	0.05	household	0.015	has	0.015	minor	0.06
this	0.025	shell	0.005	are	0.005	petty	0.04

Tabelle 1: Lexikalische Übersetzungswahrscheinlichkeiten

Gegeben die Tabelle(n) 1 mit lexikalischen Übersetzungswahrscheinlichkeiten, berechnen Sie die Übersetzungswahrscheinlichkeit nach IBM-Modell 1 für die folgenden Übersetzungen des deutschen Satzes *das Haus ist klein*.

1. the house is small
2. the house is little
3. small house the is
4. the

Nehmen Sie jeweils die wahrscheinlichste Alinierung an.

- (a) Ist das IBM-Modell 1 selbst als Übersetzungsmodell ein gutes Modell um die beste Übersetzung zu finden?
- (b) Erklären Sie, wie das Noisy-Channel-Modell manche Probleme des IBM-Modell-1-Übersetzungsmodells behebt.