

Maschinelles Übersetzen natürlicher Sprachen

Aufgabe 1

Es seien $V_E = \{\text{children, let's, play}\}$ und $V_F = \{\text{Kinder, lasst, spielen, uns}\}$. Des weiteren sei ein bilingualer Korpus $d = (f_1, e_1)(f_2, e_2)$ gegeben, wobei

$$(f_1, e_1) = (\text{Kinder spielen, children play})$$

$$(f_2, e_2) = (\text{lasst uns spielen, let's play})$$

Trainieren Sie mit dem in der Vorlesung vorgestellten Algorithmus ausgehend vom Korpus d ein Wörterbuch $t: V_E \times V_F \rightarrow [0, 1]$. Beginnen Sie mit gleichverteilten Übersetzungswahrscheinlichkeiten.

Aufgabe 2

Es seien $V_E = \{\text{du, su, ur, fur, mu}\}$ und $V_F = \{\text{kra, ban, las, gha, ra}\}$. Ferner sei das durch folgende Übersetzungswahrscheinlichkeiten bestimmte Wörterbuch t gegeben:

$t(f e)$	f				
	kra	ban	las	gha	ra
du	1	0	0	0	0
su	0	1	0	0	0
e ur	0	0	1	0	0
fur	0	0	0	1	0
mu	0	0	0	0	1

Das Längenmodell ε , das wir verwenden, ist definiert durch

$$\varepsilon(m | l) = \begin{cases} 1 & \text{falls } m = l, \\ 0 & \text{sonst.} \end{cases}$$

Übersetzen Sie den Satz "kra ban las gha ra" unter Verwendung des folgenden Bigramm-Modells b :

$b(w_2 w_1)$	w_2					
	du	su	ur	fur	mu	#
du	0	0.2	0.1	0.3	0.2	0.2
su	0.3	0	0	0.3	0.3	0.1
w ₁ ur	0.1	0	0	0.7	0	0.2
fur	0.3	0.2	0.2	0	0	0.3
mu	0.2	0.5	0	0	0	0.3
#	1	0	0	0	0	0

Dabei ist # das Stoppsymbol des Bigramm-Modells.