Aufgabe.1:

Wir bekommen immer dieselbe falsche Übersetzung für jedes Satz.

Aufgabe.2:

Diese TMX Datei speichert Übersetzungsdatei und dient zum Datenaustausch zwischen unterschiedliche Programme.

Um die Übersetzungsrichtungen zu ändern, kann man den Tag von Sprachen Z.B. <srclang> und <tuv lang='DE-DE'> verändern.

Es gibt schon über 300 Segmentationsregeln für Deutsch, im Vergleich zu den 22 englischen Unterstützungen, die deutsche sind viel mehr detailliert und umfassend.

Aufgabe.3

Die Bigramme von den Trainingskorpus:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| e(i-1)  ei | START | the | house | is | small | little | building | home | END |
| START |  |  |  |  |  |  |  |  |  |
| the | 4 |  |  |  |  |  |  |  |  |
| House |  | 2 |  |  |  |  |  |  |  |
| Is |  |  | 2 |  |  |  | 2 |  |  |
| Small |  |  |  | 3 |  |  |  |  |  |
| Little |  |  |  | 1 |  |  |  |  |  |
| Building |  | 1 |  |  |  |  |  | 1 |  |
| home |  | 1 |  |  |  |  |  |  |  |
| END |  |  |  |  | 3 | 1 |  |  |  |

P(e|e’)=Count(e‘,e)/Count(e), deshalb kriegen wir folgende Parameter:

P(the|START)=1, P(house|the)=0.5, P(is|house)=1, P(small|is)=0.75, P(END|small)=1, P(building|the)=0.25, P(is|building)=1, P(home|the)=0.25, P(building|home)=1, P(little|is)=0.25, P(END|little)=1.

f=’das Haus is klein’, e1=’the building is little‘, e2=‘the house is small’.

1. p(e1,f):

P(e1)= P(the|START)\* P(building|the)\* P(is|building)\* P(little|is)\* P(END|little)

= 1\*0.25\*1\*0.25\*1= 0.0625

p(f|e1)=P(das|the) P(Haus|building)P(ist|is)P(klein|little)/(4+1)4

=0.4\*0.3\*0.7\*0.7/625=0.00009408

Also p(e1,f) = p(e1) p(f|e1)=0.00000588.

1. p(e2,f):

P(e2)= P(the|START)\* P(house|the)\* P(is|house)\* P(small|is)\* P(END|small)

= 1\*0.5\*1\*0.75\*1=0.375

p(f|e2)=P(das|the) P(Haus|house)P(ist|is)P(klein|small)/(4+1)4

= 0.4\*0.5\*0.7\*0.8/625=0.0001792

Also p(e2,f) = p(e2) p(f|e2)=0.0000672