

Übung 7

Evaluierung von maschinellen Übersetzungen

Die folgenden Übersetzungen zweier MÜ-Systeme und eine Referenzübersetzung sind gegeben:

- SYSTEM A: Israeli officials responsibility of airport safety
- SYSTEM B: airport security Israeli officials are responsible
- REFERENCE: Israeli officials are responsible for airport security

		Israeli	officials	responsibility	of	airport	safety
	0	1	2	3	4	5	6
Israeli	1	0	1	2	3	4	5
officials	2	1	0	1	2	3	4
are	3	2	1	1	2	3	4
responsible	4	3	2	2	2	3	4
for	5	4	3	3	3	3	4
airport	6	5	4	4	4	3	4
security	7	6	5	5	5	4	4

		airport	security	Israeli	officials	are	responsible
	0	1	2	3	4	5	6
Israeli	1	1	2	2	3	4	4
officials	2	2	2	3	2	3	4
are	3	3	3	3	3	2	3
responsible	4	4	4	4	4	3	2
for	5	5	5	5	5	4	3
airport	6	5	6	6	6	5	4
security	7	6	5	6	7	6	5

Berechnen Sie die **Wortfehlerrate** WER für beide Systeme. Dazu müssen Sie zunächst den Editierabstand berechnen, welcher die minimale Zahl an Wortersetzungen, -löschungen, und -einfügungen ist, um die Übersetzung in die Referenzübersetzung umzuwandeln.

Sie gehen folgendermaßen vor:

Berechnen Sie den Editierabstand, indem Sie die rechte Matrix auf der nächsten Seite ausfüllen, welche die Wörter der Ausgabe von SYSTEM B auf der einen Achse und die Referenzwörter auf der anderen Achse zeigt. Die Matrix links daneben zeigt die fertige Berechnung für die Ausgabe des Systems A.

Der Editierabstand $d(i, k)$ für zwei Sätze **a** und **b** wird rekursiv berechnet nach der Formel:

$$d(i, k) = \min(d(i-1, k) + 1, d(i, k-1) + 1, d(i-1, k-1) + 1_{a_i \neq b_k})$$

Dabei gilt, dass $d(i, k)$ den Wert unendlich hat, falls $i < 0$ oder $k < 0$. und $1_{a_i \neq b_k} = 1$ falls $a_i \neq b_k$, sonst 0.

Mit dem Editierabstand berechnen Sie dann die Wortfehlerrate WER beider Systeme nach der Formel:

$$WER = \frac{d(|a|, |b|)}{|b|}$$

wobei b die Referenzübersetzung ist.

Was beobachten Sie?

Nun berechnen Sie die **positionsunabhängige Wortfehlerrate PER**

Was beobachten Sie hier?

Zum Schluss berechnen Sie den **Bleu-Score** beider Sätze.

Der BLEU-Score kombiniert n -Gramm-Precision-Werte mit dem *Längenbonus* $\min(1, e^{1-\frac{r}{c}})$ nach der Formel:

$$\text{BLEU-n} = \min\left(1, e^{1-\frac{r}{c}}\right) \sqrt[n]{\prod_{i=1}^n \text{precision}_i}$$

wobei r die Länge der Referenzübersetzung und c die Länge der maschinellen Übersetzung ist.

Berechnen Sie den BLEU-Score mit n-Grammen der Längen 1 bis 4.

Metric	SYSTEM A	SYSTEM B
precision (1-gram)		
precision (2-gram)		
precision (3-gram)		
precision (4-gram)		
brevity penalty		
BLEU-4		