

## Übung 4

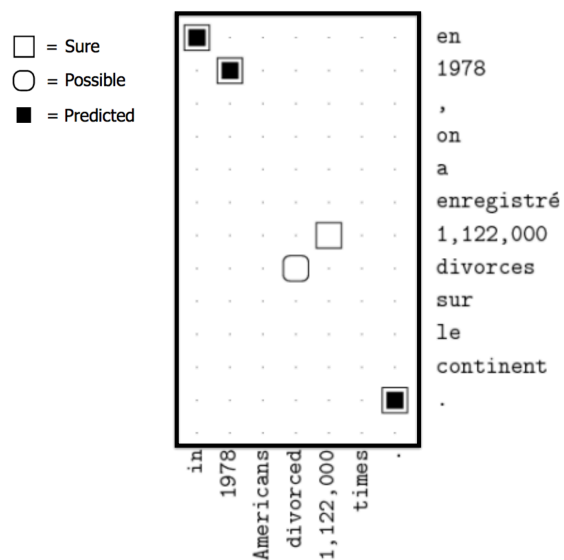
### Evaluation von Wortalignierungen

Die Alignierungsfehlerrate (AER) wird nach folgender Formel berechnet:

$$AER(A, S, P) = 1 - \frac{|A \cap S| + |A \cap P|}{|A| + |S|} \quad (1)$$

mit  $P(A, S) = \frac{|A \cap S|}{|A|}$  und  $R(A, S) = \frac{|A \cap S|}{|S|}$ , wobei  $S$  sichere Alignierungen,  $P$  mögliche Alignierungen und  $A$  automatisch erzeugte Alignierungen sind.

Berechnen Sie die Alignierungsfehlerrate für die folgenden Satzpaare:



In einer Shared Task für Wortalignierung sollen zwei Systeme verglichen werden. Berechnen Sie AER, Precision, Recall, und den  $F_\alpha$ -Score für die beiden Systeme auf Basis folgender Daten:

	SYSTEM 1	SYSTEM 2
$ S $	100	100
$ A $	100	100
$ A \cap S $	25	50
$ A \cap P $	75	50

Definition des  $F_\alpha$ -Scores:

$$F_\alpha - \text{SCORE}(A, S, \alpha) = \frac{1}{\frac{\alpha}{P} + \frac{1-\alpha}{R}}$$

Was beobachten Sie?

### Symmetrisierung von Wortalignierungen

Gegeben die Sätze  $e$  und  $f$  mit  $l_e = 4$  und  $l_f = 5$  und die beiden IBM Modell Alignments  $e2f$  (rechts) und  $f2e$  (links):

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$e_1$					
$e_2$					
$e_3$					
$e_4$					

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$e_1$					
$e_2$					
$e_3$					
$e_4$					

Berechnen Sie die symmetrisierten Wortalignierungen gemäß dem folgenden Pseudocode von den Vorlesungsfolien:

**grow\_diag\_final(e2f, f2e)**

neighbouring =  $\{(-1,0),(0,-1),(1,0),(0,1),(-1,-1),(-1,1),(1,-1),(1,1)\}$

alignment  $A$  = intersect(e2f,f2e)

**grow\_diag(); final(e2f); final(f2e)**

**grow\_diag()**

**while** new points added **do**

**for all** English words  $e \in [1...e_n]$ , foreign words  $f \in [1...f_n]$ ,  $(e, f) \in A$  **do**

**for all** neighbouring alignment points  $(e_{new}, f_{new})$  **do**

**if**  $(e_{new}$  unaligned **or**  $f_{new}$  unaligned) **and**  $(e_{new}, f_{new}) \in \text{union}(e2f, f2e)$  **then**  
add  $(e_{new}, f_{new})$  to  $A$

**final(a)**

**for all** English words  $e_{new} \in [1...e_n]$ , foreign words  $f_{new} \in [1...f_n]$  **do**

**if**  $(e_{new}$  unaligned **or**  $f_{new}$  unaligned) **and**  $(e_{new}, f_{new}) \in a$  **then**  
add  $(e_{new}, f_{new})$  to  $A$

Schritte:

- Berechnen Sie die Schnittmenge der Alignierungen.
- Wenden Sie die `grow-diag()`-Methode an und fügen Sie Nachbarpunkte hinzu, die in der Vereinigungsmenge aber nicht in der Schnittmenge enthalten sind.
- Wenden Sie die `final()`-Methode an, um weitere noch nicht alignierte Wörter zu alignieren.

Nun seien die beiden Sätze  $e$  und  $f$  der Länge 3 gegeben, und die Alignments  $A_{ef} : 1 \rightarrow 1, 2 \rightarrow 3$  und  $A_{fe} : 1 \rightarrow 1, 2 \rightarrow 1, 2 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 2$ :

- Folgen Sie exakt dem Pseudocode oben, um die Symmetrisierung der Alignierungen zu berechnen.
- Ist die `grow-diag()`-Methode deterministisch?