

# Statistical Machine Translation

-word based models-

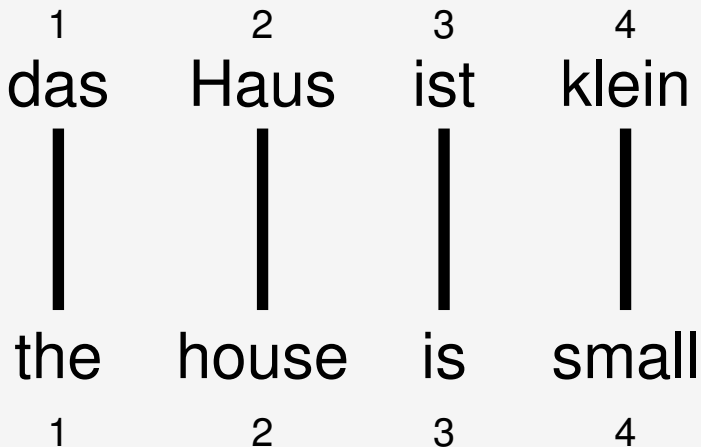
**Artem Sokolov**

Computerlinguistik

Universität Heidelberg

Sommersemester 2015

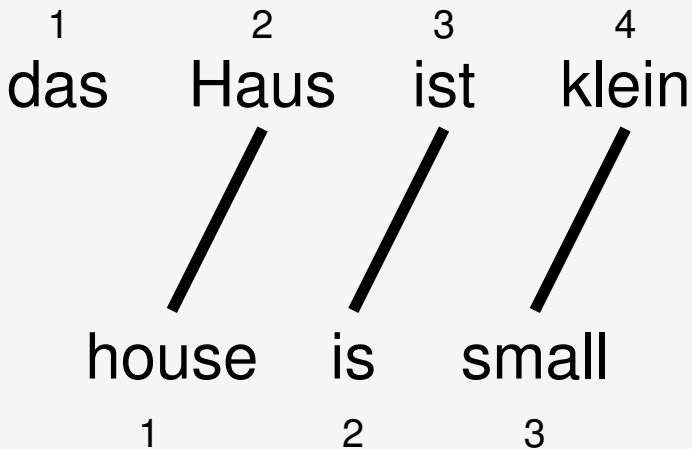
material from P. Koehn, S. Riezler, M. Schlesinger



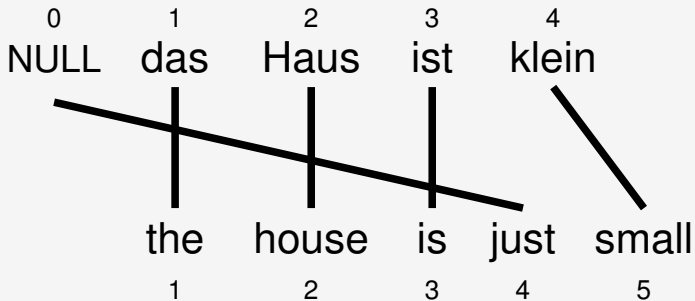
1	2	3	4
klein	ist	das	Haus
the	house	is	small
1	2	3	4

The diagram illustrates the reordering of German words to their English equivalents. The German words are arranged in a row: 'klein', 'ist', 'das', 'Haus'. The English words are arranged in a row below: 'the', 'house', 'is', 'small'. Thick black lines connect the German words to the English words: 'klein' connects to 'small', 'ist' connects to 'is', 'das' connects to 'the', and 'Haus' connects to 'house'. The numbers 1, 2, 3, and 4 are placed above and below the words to indicate their original positions.





1                    2                    3                    4                    5  
das            Haus            ist            ja            klein  
  \                    \                    \                    /  
the            house            is            small  
1                    2                    3                    4

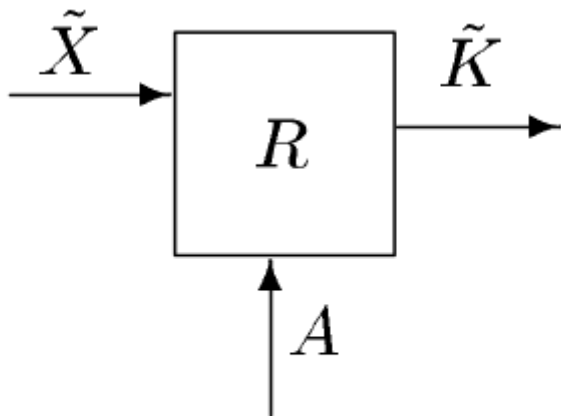


1	2	3	4
das	Haus	ist	klein
the	house	is	small
1	2	3	4

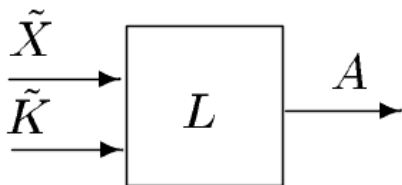


das		Haus		ist		klein	
$e$	$t(e f)$	$e$	$t(e f)$	$e$	$t(e f)$	$e$	$t(e f)$
the	0.7	house	0.8	is	0.8	small	0.4
that	0.15	building	0.16	's	0.16	little	0.4
which	0.075	home	0.02	exists	0.02	short	0.1
who	0.05	household	0.015	has	0.015	minor	0.06
this	0.025	shell	0.005	are	0.005	petty	0.04

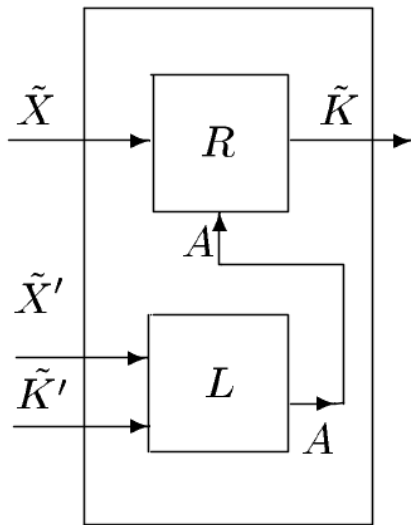
$$\begin{aligned}
 p(e, a|f) &= \frac{\epsilon}{5^4} \times t(\text{the}|\text{das}) \times t(\text{house}|\text{Haus}) \times t(\text{is}|\text{ist}) \times t(\text{small}|\text{klein}) \\
 &= \frac{\epsilon}{5^4} \times 0.7 \times 0.8 \times 0.8 \times 0.4 \\
 &= 0.0029\epsilon
 \end{aligned}$$



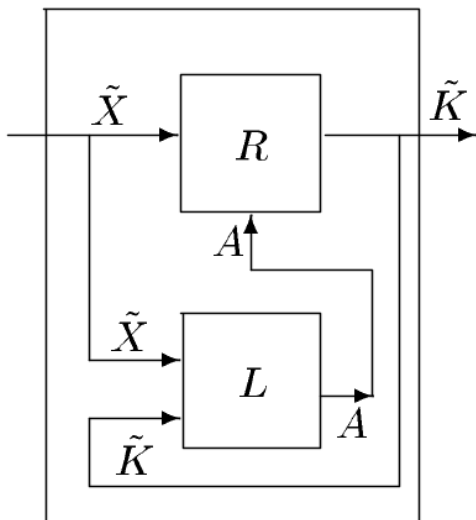
(a) Varying classifier.



(b) Learning as an estimate of parameter  $A$ .

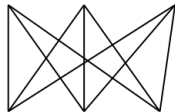


(c) Learning classifier.



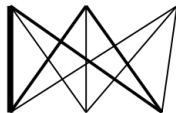
(d) Classifier based on unsupervised learning.

... la maison ... la maison bleue... la fleur ...



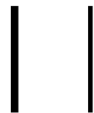
... the house ... the blue house ... the flower ...

... la maison ... la maison bleue... la fleur ...



... the house ... the blue house ... the flower ...

... la maison ... la maison bleue... la fleur ...



... the house ... the blue house ... the flower ...



... la maison ... la maison bleue... la fleur ...



... the house ... the blue house ... the flower ...

## EM-Algorithm

... la maison ... la maison bleue... la fleur ...  
  
 ... the house ... the blue house ... the flower ...

$$p(\text{la}|\text{the}) = 0.453$$

$$p(\text{le}|\text{the}) = 0.334$$


$$p(\text{maison}|\text{house}) = 0.876$$


$$p(\text{bleue}|\text{blue}) = 0.563$$


das Haus  
the house

das Buch  
the book

ein Buch  
a book

das Haus  
  
 the house

das Buch  
  
 the book

ein Buch  
  
 a book

<i>e</i>	<i>f</i>	initial	1st it.	2nd it.	3rd it.	...	final
the	das	0.25	0.5	0.6667	0.7996	...	1
book	das	0.25	0.25	0.1795	0.1115	...	0
house	das	0.25	0.25	0.1538	0.0889	...	0
the	buch	0.25	0.25	0.1795	0.1115	...	0
book	buch	0.25	0.5	0.6667	0.7996	...	1
a	buch	0.25	0.25	0.1538	0.0889	...	0
book	ein	0.25	0.5	0.5	0.5	...	0.5
a	ein	0.25	0.5	0.5	0.5	...	0.5
the	haus	0.25	0.5	0.5	0.5	...	0.5
house	haus	0.25	0.5	0.5	0.5	...	0.5