Lexical Translation and Alignment

Dr. Mariana Neves
(adapted from the original slides
of Prof. Philipp Koehn)

November 7th, 2016

Word-based models

- Statistical machine translation models developed in the IBM Candice project between late 1980s and early 1990s
- Not state of art approaches anymore, but they are still current today

Overview

- Lexical translation
- Alignment

Overview

- **Lexical translation**
- Alignment

Lexical Translation

- How to translate a word → look up in dictionary

  Haus — house, building, home, household, shell.

- Note: In all lectures, we translate from a foreign language into English
- Multiple translations
  - some more frequent than others
  - for instance: house, and building most common
  - special cases: Haus of a snail is its shell
- How can we learn about word frequencies?

Collect Statistics

Look at a parallel corpus (German text along with English translation)

Translation of Haus	Count
house	8,000
building	1,600
home	200
household	150
shell	50

- What is the problem of this approach?

Collect Statistics

Drawbacks of word-based statistics

- The context in which the word appears is ignored
- The size of the dataset influences in the statistics
- The domain or the style of the dataset influence the statistics

Estimate Translation Probabilities

Maximum likelihood estimation

$$p_f(e) = \begin{cases} 0.8 & \text{if } e = \text{house}, \\ 0.16 & \text{if } e = \text{building}, \\ 0.02 & \text{if } e = \text{home}, \\ 0.015 & \text{if } e = \text{household}, \\ 0.005 & \text{if } e = \text{shell}. \end{cases}$$

For a foreign word f, $p_f(e)$ indicates the probability of the English word e to be a translation for f.

Estimate Translation Probabilities

Properties of the probability distribution:

$$p_f(e) = \begin{cases} 0.8 & \text{if } e = \text{house}, \\ 0.16 & \text{if } e = \text{building}, \\ 0.02 & \text{if } e = \text{home}, \\ 0.015 & \text{if } e = \text{household}, \\ 0.005 & \text{if } e = \text{shell}. \end{cases}$$

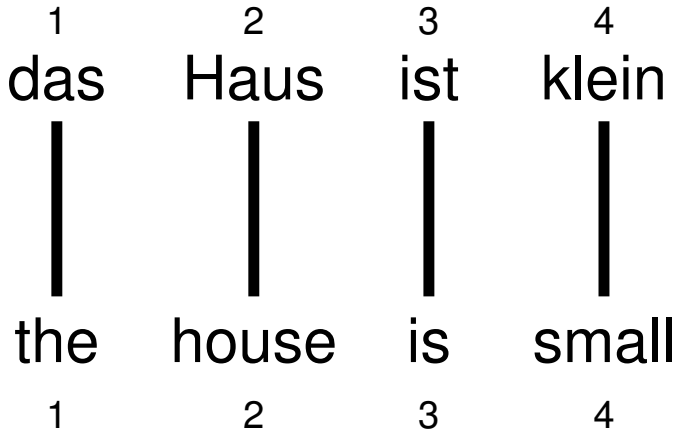
$$\sum_e p_f = 1 \qquad \forall e : 0 \leq p_f \leq 1$$

Overview

- Lexical translation
- **Alignment**

Alignment

- In a parallel text (or when we translate), we align words in one language with the words in the other



- Word positions are numbered 1–4

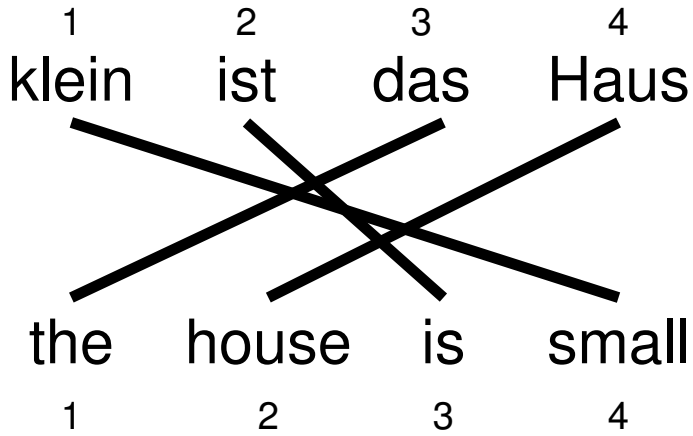
Alignment Function

- Formalizing alignment with an alignment function
- Mapping an English target word at position i to a German source word at position j with a function $a : i \rightarrow j$
- Example

$$a : \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3, 4 \rightarrow 4\}$$

Reordering

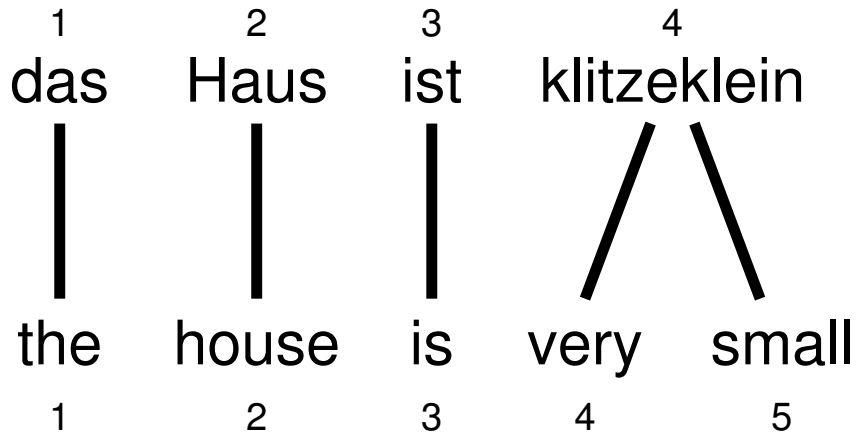
Words may be reordered during translation



$$a : \{1 \rightarrow 3, 2 \rightarrow 4, 3 \rightarrow 2, 4 \rightarrow 1\}$$

One-to-Many Translation

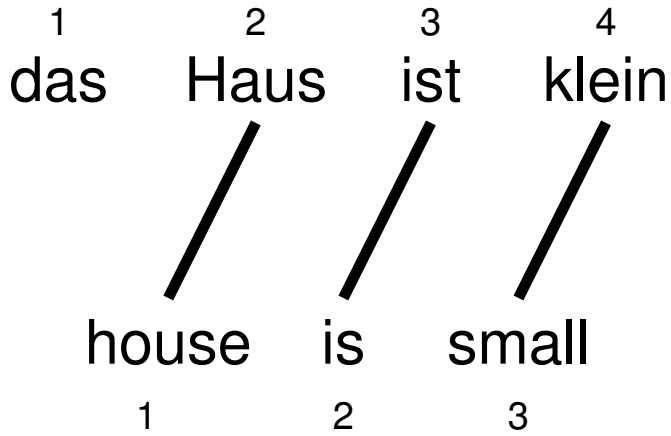
A source word may translate into multiple target words



$$a : \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3, 4 \rightarrow 4, 5 \rightarrow 4\}$$

Dropping Words

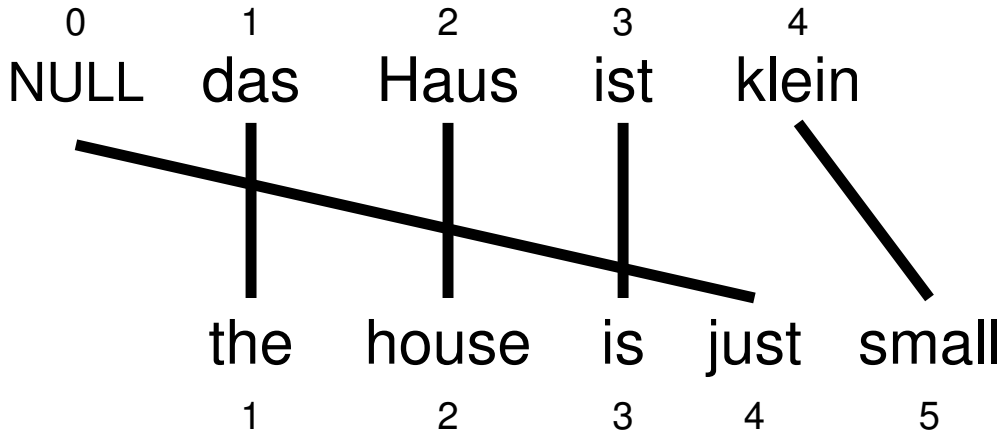
Words may be dropped when translated
(German article das is dropped)



$a : \{1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 4\}$

Inserting Words

- Words may be added during translation
  - The English just does not have an equivalent in German
  - We still need to map it to something: special NULL token



$$a : \{1 \to 1, 2 \to 2, 3 \to 3, 4 \to 0, 5 \to 4\}$$

Alignment Model

- It allows dropping and adding words
- Translation of a foreign language (input) f to English (output) e
- Each output word is linked to only one input word (including the NULL token)
- But an input can be linked to none or many output word (e.g., klitzeklein)

Suggested reading

- Statistical Machine Translation, Philipp Koehn (section 4.1).