

CA446

Statistical Machine Translation



# Week 7: Phrase-based Translation Model

Lecturer: Qun Liu

Lab Tutor: Xiaofeng Wu, Iacer Calixto

2<sup>nd</sup> Semester, 2014-2015 Academic Year

<http://computing.dcu.ie/~qliu/CA446>

# Content

## Phrase-based Translation Model

Learning a Phrase Translation Table

Bidirectional Word Alignment

Phrase Pair Extraction

Phrase Translation Probability

Exercises

# Phrase-based translation models

- Word-based models translate **words** as atomic units.

# Phrase-based translation models

- Word-based models translate **words** as atomic units.
- Phrase-based models translate **phrases** as atomic units.

# Phrase-based translation models

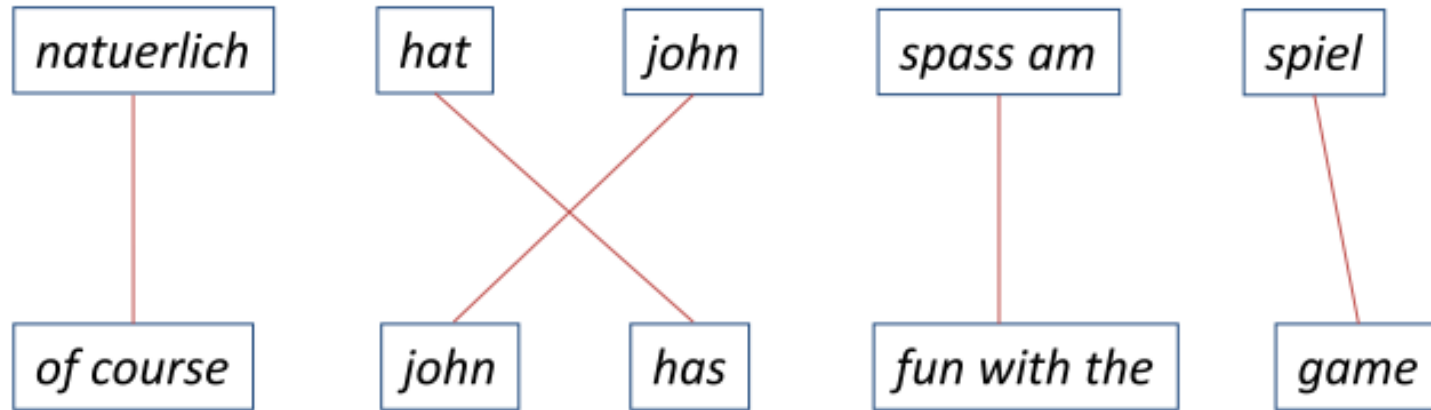
- Word-based models translate **words** as atomic units.
- Phrase-based models translate **phrases** as atomic units.
- A phrase is a contiguous sequence of words in a sentence.

# Phrase-based translation models

- Word-based models translate **words** as atomic units.
- Phrase-based models translate **phrases** as atomic units.
- A phrase is a contiguous sequence of words in a sentence.

Phrase-based models are the “standard” model in statistical machine translation.

# Phrase-based translation



- Source sentence is segmented into phrases.
- Each phrase is translated into target language.
- Phrases are re-ordered.

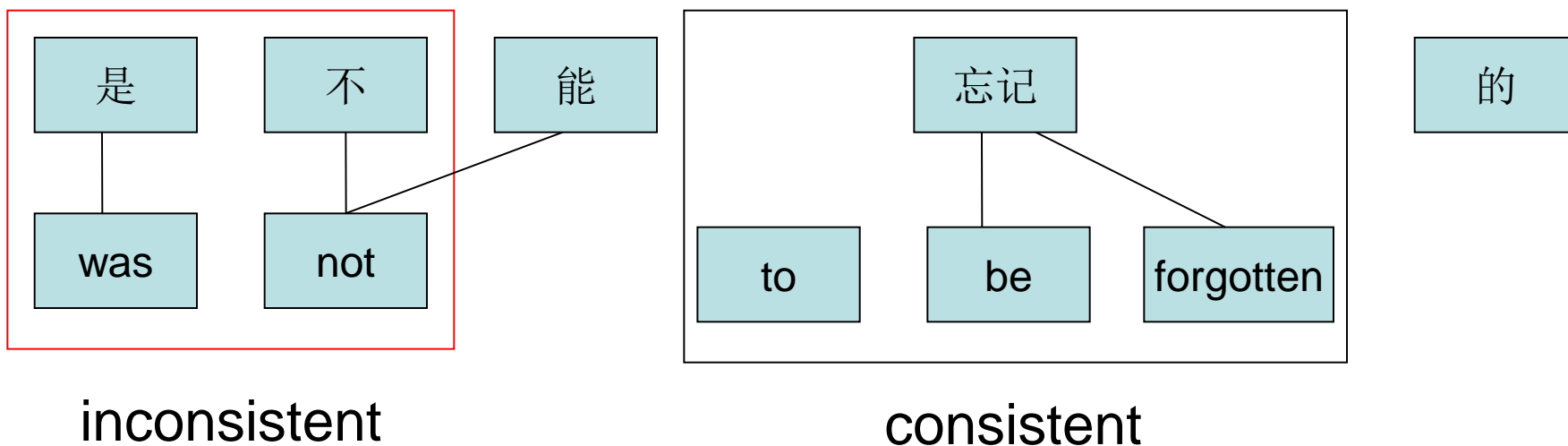
# Phrase-based translation



- A monolingual phrase:
  - A phrase can be **any contiguous sequence of words** in a sentence
  - A phrase is not necessarily syntactic well-formed
  - A phrase is not necessarily semantically meaningful
- A bilingual phrase pair should be consistent with word alignment.



# Bilingual Phrase Pairs



# Advantages of phrase-based translation



1. Can handle non-compositional phrases or **idioms** (e.g. *shooting the breeze*)

# Advantages of phrase-based translation



1. Can handle non-compositional phrases or **idioms** (e.g. *shooting the breeze*)
2. Takes **context** into account (e.g. *I go* translated as *Ich gehe* rather than *Ich gehen*)

# Advantages of phrase-based translation



1. Can handle non-compositional phrases or **idioms** (e.g. *shooting the breeze*)
2. Takes **context** into account (e.g. *I go* translated as *Ich gehe* rather than *Ich gehen*)
3. The more data we have the **longer phrases** we can learn (*Nice to meet you, Can I have the bill please?*)

# Phrase translation table

- Main knowledge source: table with phrase translations and their probabilities
- Example: phrase translations for **natuerlich**

Translation	Probability
of course	0.5
naturally	0.3
of course ,	0.15
, of course ,	0.05

# Phrase translation table

- Real example taken from Europarl for the German phrase **den Vorschlag**

English	Probability	English	Probability
the proposal	0.6277	the suggestions	0.0114
's proposal	0.1068	the proposed	0.0114
a proposal	0.0341	the motion	0.0091
the idea	0.025	the idea of	0.0091
this proposal	0.0227	the proposal ,	0.0068
proposal	0.0205	its proposal	0.0068
of the proposals	0.0159	it	0.0068
the proposals	0.0159	.....	

# Content

- Phrase-based Translation Model
- **Learning a Phrase Translation Table**
- Bidirectional Word Alignment
- Phrase Pair Extraction
- Phrase Translation Probability
- Exercises

# Learning a phrase translation table

**Task:** learn the model from a parallel corpus



# Learning a phrase translation table

**Task:** learn the model from a parallel corpus

**Three stages:**

# Learning a phrase translation table

**Task:** learn the model from a parallel corpus

**Three stages:**

1. word alignment: using IBM models or other method

# Learning a phrase translation table

**Task:** learn the model from a parallel corpus

**Three stages:**

1. word alignment: using IBM models or other method
2. extraction of phrase pairs

# Learning a phrase translation table

**Task:** learn the model from a parallel corpus

**Three stages:**

1. word alignment: using IBM models or other method
2. extraction of phrase pairs
3. scoring phrase pairs

# Content

- Phrase-based Translation Model
- Learning a Phrase Translation Table
- **Bidirectional Word Alignment**
- Phrase Pair Extraction
- Phrase Translation Probability
- Exercises

# A problem with word alignment and IBM models

Each target word can be aligned to at most one source word. Therefore, it's not possible to end up with an alignment of one target word to many source words

herzlichen glückwunsch



congratulations



# How to fix this?

- Compute word alignments in both directions!

*congratulations*  
/ \  
*herzlichen glückwunsch*



- In this way, we can get many-to-one alignments as well as one-to-many alignments.

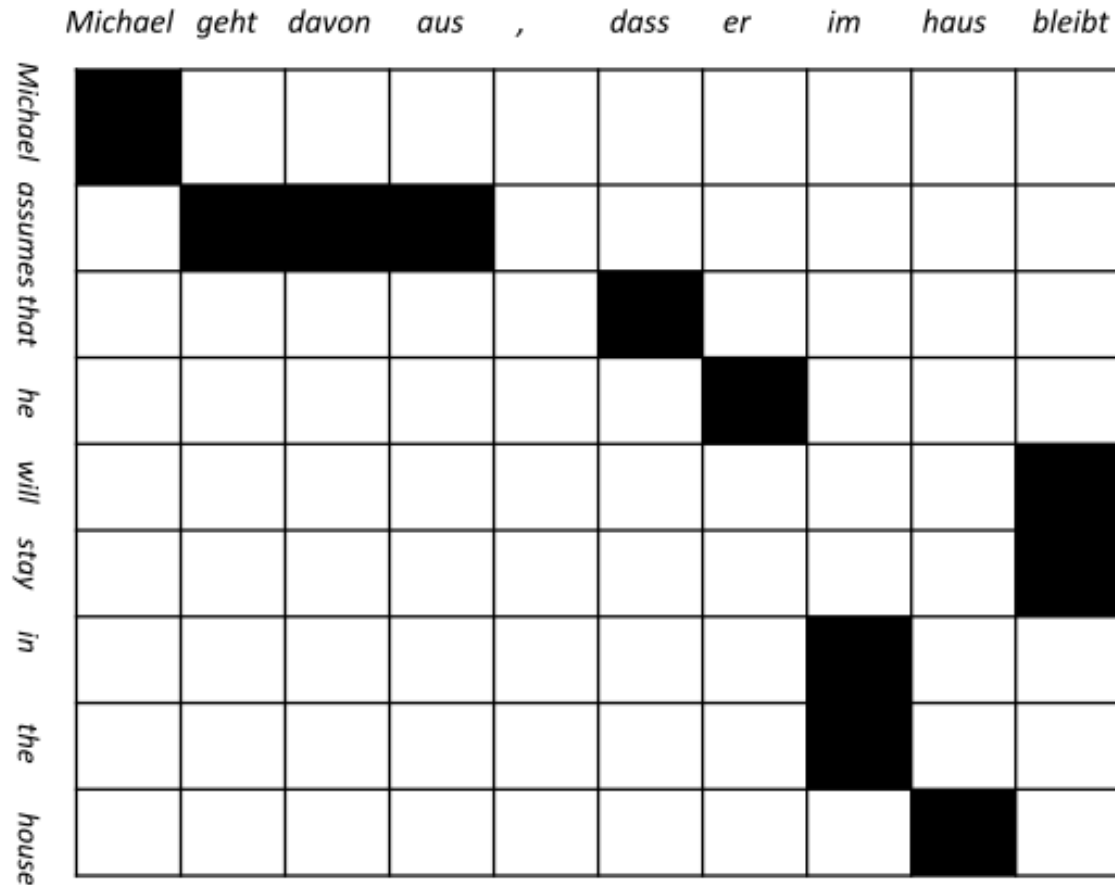
# Bidirectional word alignment

Algorithm of Bidirectional word alignment:

1. Using IBM Models to do word alignment in one direction.
2. Using IBM Models to do word alignment in the other direction.
3. Merge the above two alignments.



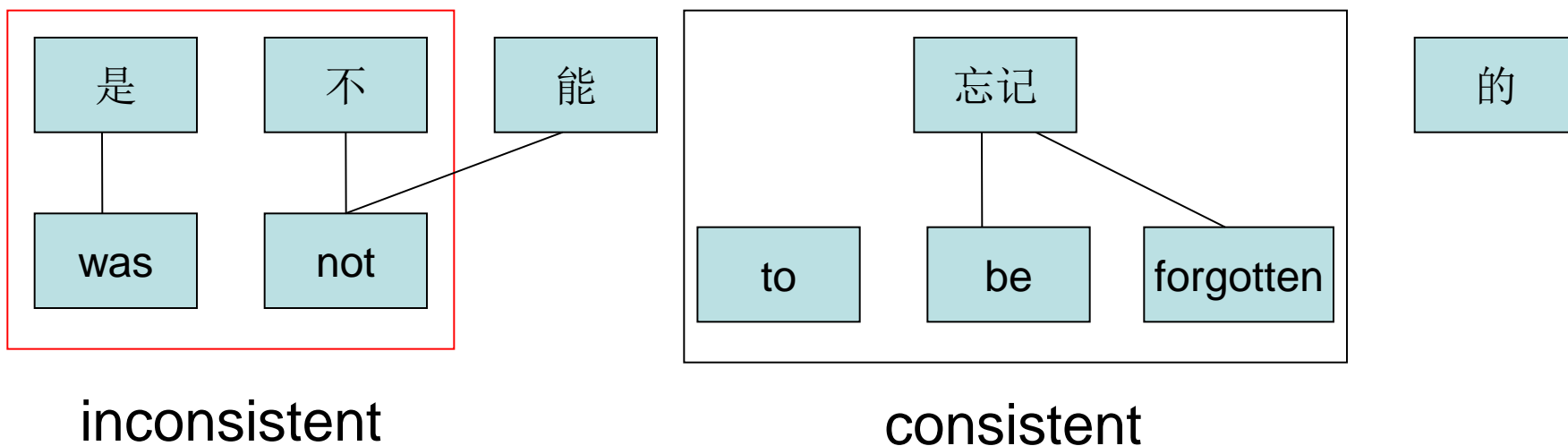
# Bidirectional word alignment



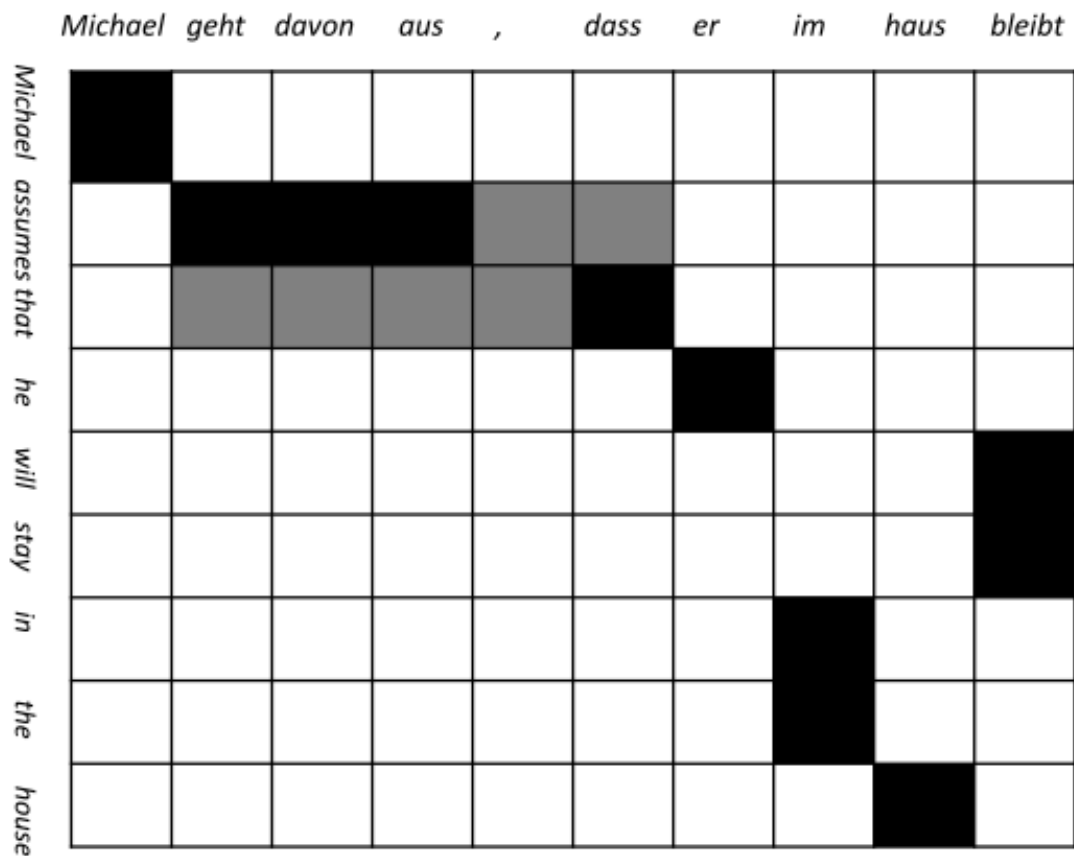
# Content

- Phrase-based Translation Model
- Learning a Phrase Translation Table
- Bidirectional Word Alignment
- **Phrase Pair Extraction**
- Phrase Translation Probability
- Exercises

# Bilingual Phrase Pairs



# Extracting phrase pairs



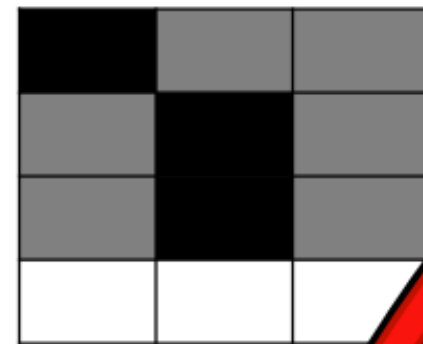
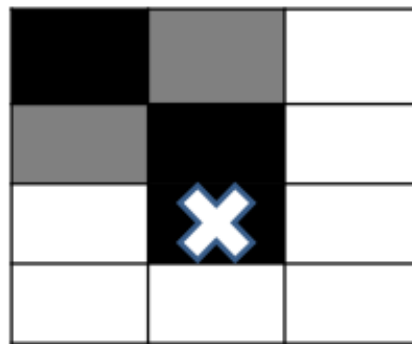
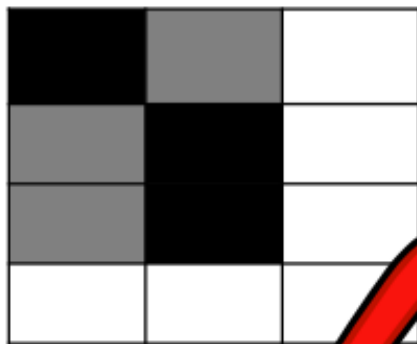
extract phrase pair consistent with word alignment:  
 assumes that / geht davon aus , dass

# Consistent with word alignment

A phrase pair  $(e, f)$  is consistent with a bidirectional word alignment  $A$  if and only if

- For all words  $e_i$  in  $e$ , if  $e_i$  is aligned to a word  $f_j$  in  $A$ , then  $f_j$  is in  $f$ .
- For all word  $f_j$  in  $f$ , if  $f_j$  is aligned to a word  $e_i$  in  $A$ , then  $e_i$  is in  $e$ .
- There exists  $e_i$  in  $e$ ,  $f_j$  in  $f$  :  $(e_i, f_j)$  in  $A$

# Consistent with word alignment



# Phrase Pair Extraction

*Michael geht davon aus , dass er im haus bleibt*

<i>Michael</i>	█								
<i>assumes that</i>		█	█	█					
<i>he</i>					█				
<i>will</i>						█			
<i>stay</i>									█
<i>in</i>							█		
<i>the</i>							█		
<i>house</i>								█	

# Phrase Pair Extraction



michael | michael

assumes | geht davon aus / geht davon aus ,

that | dass / , dass

he | er

will stay | bleibt

in the | im

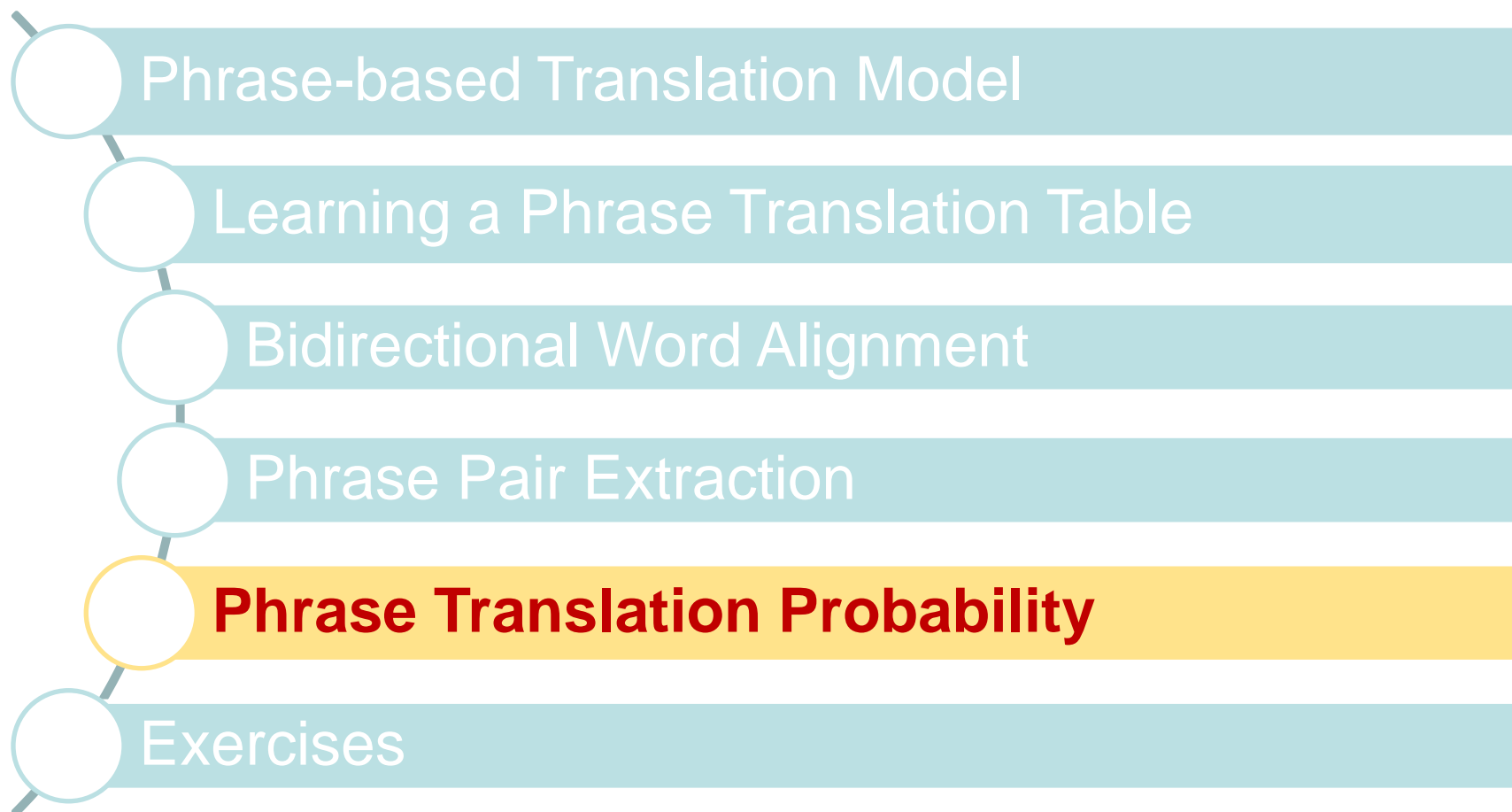
house | haus



# Phrase Pair Extraction

- michael assumes | michael geht davon aus / michael geht davon aus ,
- assumes that | geht davon aus , dass
- assumes that he | geht davon aus , dass er
- that he | dass er / , dass er
- in the house | im haus
- michael assumes that | michael geht davon aus , dass
- michael assumes that he | michael geht davon aus , dass er
- michael assumes that he will stay in the house | michael geht davon aus ,  
dass er im haus bleibt
- assumes that he will stay in the house | geht davon aus , dass er im haus  
bleibt
- that he will stay in the house | dass er im haus bleibt / dass er im haus bleibt ,
- he will stay in the house | er im haus bleibt
- will stay in the house | im haus bleibt

# Content



# Scoring Phrase Translations

- Phrase pair extraction: collect all phrase pairs from the data
- Phrase pair scoring: assign probabilities to phrase translations
- Score by relative frequency:

$$\phi(f|e) = \frac{\textit{count}(e, f)}{\sum_{f_i} \textit{count}(e, f_i)}$$

# Content

- Phrase-based Translation Model
- Learning a Phrase Translation Table
- Bidirectional Word Alignment
- Phrase Pair Extraction
- Phrase Translation Probability
- Exercises**

# Exercises

- List all phrase pairs that are consistent with the following word alignment:

	<i>A</i>	<i>B</i>	<i>C</i>
<i>x</i>	■	□	□
<i>y</i>	□	■	□
<i>z</i>	□	□	■

# Exercises

- List all phrase pairs that are consistent with the following word alignment:

	A	B	C
X	■		
Y			
Z			■

# Exercises

- Given the following sentence alignments:  
The book                      blue book  
An leabhar                    leabhar gorm
- State what the following translation probabilities will be after two iterations of the Expectation Maximisation algorithm and show all the steps you followed to arrive at these values:

i. $t(\text{leabhar} \text{The})$	ii. $t(\text{An} \text{The})$	iii. $t(\text{gorm} \text{The})$
iv. $t(\text{leabhar} \text{book})$	v. $t(\text{An} \text{book})$	vi. $t(\text{gorm} \text{book})$
vii. $t(\text{leabhar} \text{blue})$	viii. $t(\text{An} \text{blue})$	ix. $t(\text{gorm} \text{blue})$

# Questions

- What is the difference between a phrase-based translation model and a word-based translation model?
- List two advantages of phrase-based Models.
- Why do we still care about word-based models?





# Discussion

# Acknowledgement



Parts of the content of this lecture are taken from previous lectures and presentations given by Jennifer Foster, Declan Groves, Yvette Graham, Kevin Knight, Josef van Genabith, Andy Way.