# Introduction to Phrase-based Statistical Machine Translation

Fabienne Cap

UPPSALA UNIVERSITY SWEDEN

# Where are we?

| Type | Date | Time | Place | Topic | Reading / Assignments |
|---|---|---|---|---|---|
| F | 2016-03-30 | 10-12 | 6-K1031 | Introduction (SS) | Koehn 1; JM 25.1-2; Hutchins; CFMF |
| F | 2016-03-30 | 14-16 | 6-K1031 | MT evaluation (SS) | Koehn 8; JM 25.9 |
| F | 2016-04-04 | 10-12 | 2-0076 | MT in practice (Convertus) - guest lecture | |
| L | 2016-04-06 | 10-12 | Chomsky | MT in practice (AS) | lab report 1 |
| F | 2016-04-11 | 10-12 | 2-0076 | Introduction to SMT (FC) | Koehn Ch 4, Ch 7, KK97 |
| L | 2016-04-13 | 10-12 | Chomsky | Word-based SMT (SS) | lab report 2 |
| L | 2016-04-18 | 10-12 | Chomsky | Word-based SMT (SS) | lab report 2 |
| F | 2016-04-18 | 14-16 | 2-1077 | Machine translation at Semantix, a translation provider - guest lecture | |
| F | 2016-04-20 | 10-12 | 6-K1031 | Parallel Corpora, Alignment (AS) | Koehn 2-4, JT 3-4, KK97, KK99 |
| L | 2016-04-25 | 10-12 | Chomsky | Parallel corpora & alignment (AS) | lab report 3 |
| F | 2016-04-27 | 10-12 | 2-0076 | Phrase-based SMT (FC) | Koehn Ch 5 |
| L | 2016-05-02 | 10-12 | Chomsky | Phrase-based SMT (AS) | lab report 4 |
| F | 2016-05-04 | 10-12 | 6-K1031 | Decoding (CH) | Koehn Ch 6 |
| L | 2016-05-09 | 10-12 | Chomsky | Phrase-based SMT (AS) | lab report 4 |
| F | 2016-05-11 | 10-12 | 2-0076 | Tree-based SMT & MT for morphologically rich languages (SS, FC) | Koehn 10.2, 11 |
| F | 2016-05-16 | 10-12 | 2-0076 | Document-wide decoding & Neural MT (CH) | |
| L | 2016-05-18 | 10-12 | Chomsky | Document-wide decoding lab (AS) | oral lab report 5 |
| S | 2016-05-23 | 10-12 | 2-0076 | Seminar - master student presentations | |
| S | 2016-05-25 | 10-12 | 6-K1031 | Seminar - master student presentations | |

# Where are we?

| Type | Date | Time | Place | Topic | Reading / Assignments |
|---|---|---|---|---|---|
| F | 2016-03-30 | 10-12 | 6-K1031 | Introduction (SS) | Koehn 1; JM 25.1-2; Hutchins; CFMF |
| F | 2016-03-30 | 14-16 | 6-K1031 | MT evaluation (SS) | Koehn 8; JM 25.9 |
| F | 2016-04-04 | 10-12 | 2-0076 | MT in practice (Convertus) - guest lecture | |
| L | 2016-04-06 | 10-12 | Chomsky | MT in practice (AS) | lab report 1 |
| F | 2016-04-11 | 10-12 | 2-0076 | Introduction to SMT (FC) | Koehn Ch 4, Ch 7, KK97 |
| L | 2016-04-13 | 10-12 | Chomsky | Word-based SMT (SS) | lab report 2 |
| L | 2016-04-18 | 10-12 | Chomsky | Word-based SMT (SS) | lab report 2 |
| F | 2016-04-18 | 14-16 | 2-1077 | Machine translation at Semantix, a translation provider - guest lecture | |
| F | 2016-04-20 | 10-12 | 6-K1031 | Parallel Corpora, Alignment (AS) | Koehn 2-4, JT 3-4, KK97, KK99 |
| L | 2016-04-25 | 10-12 | Chomsky | Parallel corpora & alignment (AS) | lab report 3 |
| F | 2016-04-27 | 10-12 | 2-0076 | Phrase-based SMT (FC) | Koehn Ch 5 |
| L | 2016-05-02 | 10-12 | Chomsky | Phrase-based SMT (AS) | lab report 4 |
| F | 2016-05-04 | 10-12 | 6-K1031 | Decoding (CH) | Koehn Ch 6 |
| L | 2016-05-09 | 10-12 | Chomsky | Phrase-based SMT (AS) | lab report 4 |
| F | 2016-05-11 | 10-12 | 2-0076 | Tree-based SMT & MT for morphologically rich languages (SS, FC) | Koehn 10.2, 11 |
| F | 2016-05-16 | 10-12 | 2-0076 | Document-wide decoding & Neural MT (CH) | |
| L | 2016-05-18 | 10-12 | Chomsky | Document-wide decoding lab (AS) | oral lab report 5 |
| S | 2016-05-23 | 10-12 | 2-0076 | Seminar - master student presentations | |
| S | 2016-05-25 | 10-12 | 6-K1031 | Seminar - master student presentations | |

# Goals for Today

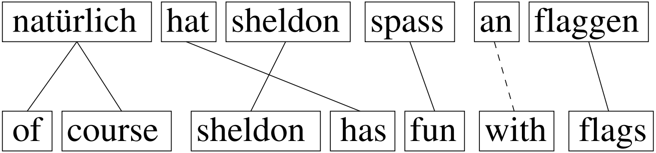
Why phrases?

Symmetrisation of word alignment

Phrase extraction and scoring

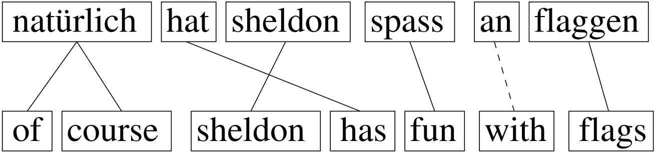
Log-Linear Model

# Word-based vs. Phrase-based SMT

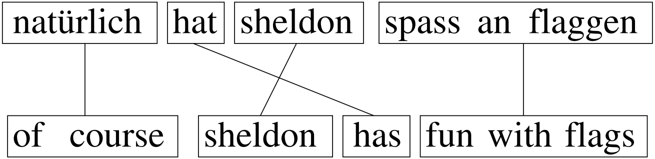



Word-based models translate **words** as atomic units

# Word-based vs. Phrase-based SMT




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




Phrase-based models translate **phrases** as atomic units

# Phrase-based SMT

- A phrase is a **continuous** sequence of words
- Not neccessarily a linguistic phrase (!):
  Spass an - fun with the
  $\rightarrow$ using only linguistic phrases **hurts** translation quality!
- State-of-the-art for many language pairs
- Used by GoogleTranslate and others

# Advantages of Phrase-based SMT

- Many-to-many translation can handle
  **non-compositional phrases:**
  *kick the bucket - ins Gras beissen* (lit. into the grass bite)
  **compounds:**
  *blädderblocksblad - flipchart paper*

- Local context can be taken into account:
  **local word order**:
  *affaires extérieure - external affairs*
  **local agreement issues:**
  *Vorlesung* **am** *Mittwoch - lecture* **on** *Wednesday*
  *Spass* **am** *Spiel - fun* **with the** *game*

# Advantages of Phrase-based SMT II

- Translating phrases helps to reduce translation ambiguities
- Phrases of arbitrary length:
  sometimes the **entire sentence** might be covered by a phrase
- Simpler model:
  no more need to explicitly model the concepts of
  **fertility, insertion** and **deletion** of words

# Real Example

Phrase translations for **begreppet** taken from EUROPARL

| English | $\phi(\bar{t}\|\bar{s})$ | English | $\phi(\bar{t}\|\bar{s})$ |
|---|---|---|---|
| the | 0.226415 | the news | 0.012816 |
| told | 0.169811 | the report | 0.008544 |
| announcement | 0.075472 | the information | 0.008544 |
| message | 0.056604 | the back | 0.004272 |
| news | 0.056604 | the suspension | 0.004272 |
| information | 0.037736 | the death | 0.004272 |
| informed | 0.037736 | this announcement | 0.002848 |
| learnt | 0.037736 | this news | 0.002136 |
| peace of mind by ensuring | 0.027778 | a message | 0.001539 |
| insight | 0.018868 | his answer | 0.000356 |
| the announcement | 0.017088 | were told | 0.000229 |
| the message | 0.012816 | the back and | 2.917e-05 |

- lexical variation (announcement, message, news, told, . . . )
- Morphological variation (information, informed)
- Included function words (the, a, were, this)
- Noise (the, the back and)

# Real Example

Phrase translations for **begreppet** taken from EUROPARL

| English | $\phi(\bar{t}\|\bar{s})$ | English | $\phi(\bar{t}\|\bar{s})$ |
|---|---|---|---|
| the | 0.226415 | the news | 0.012816 |
| **told** | 0.169811 | the report | 0.008544 |
| **announcement** | 0.075472 | the information | 0.008544 |
| **message** | 0.056604 | the back | 0.004272 |
| **news** | 0.056604 | the suspension | 0.004272 |
| information | 0.037736 | the death | 0.004272 |
| informed | 0.037736 | this announcement | 0.002848 |
| learnt | 0.037736 | this news | 0.002136 |
| peace of mind by ensuring | 0.027778 | a message | 0.001539 |
| insight | 0.018868 | his answer | 0.000356 |
| the announcement | 0.017088 | were told | 0.000229 |
| the message | 0.012816 | the back and | 2.917e-05 |

- lexical variation (announcement, message, news, told, ...)
- Morphological variation (information, informed)
- Included function words (the, a, were, this)
- Noise (the, the back and)

# Real Example

Phrase translations for **begreppet** taken from EUROPARL

| English | $\phi(\bar{t}\|\bar{s})$ | English | $\phi(\bar{t}\|\bar{s})$ |
|---|---|---|---|
| the | 0.226415 | the news | 0.012816 |
| told | 0.169811 | the report | 0.008544 |
| announcement | 0.075472 | the information | 0.008544 |
| message | 0.056604 | the back | 0.004272 |
| news | 0.056604 | the suspension | 0.004272 |
| **information** | 0.037736 | the death | 0.004272 |
| **informed** | 0.037736 | this announcement | 0.002848 |
| learnt | 0.037736 | this news | 0.002136 |
| peace of mind by ensuring | 0.027778 | a message | 0.001539 |
| insight | 0.018868 | his answer | 0.000356 |
| the announcement | 0.017088 | were told | 0.000229 |
| the message | 0.012816 | the back and | 2.917e-05 |

- lexical variation (announcement, message, news, told, . . . )
- Morphological variation (information, informed)
- Included function words (the, a, were, this)
- Noise (the, the back and)

# Real Example

Phrase translations for **begreppet** taken from EUROPARL

| English | $\phi(\bar{t}\|\bar{s})$ | English | $\phi(\bar{t}\|\bar{s})$ |
|---|---|---|---|
| the | 0.226415 | the news | 0.012816 |
| told | 0.169811 | the report | 0.008544 |
| announcement | 0.075472 | the information | 0.008544 |
| message | 0.056604 | the back | 0.004272 |
| news | 0.056604 | the suspension | 0.004272 |
| information | 0.037736 | the death | 0.004272 |
| informed | 0.037736 | **this** announcement | 0.002848 |
| learnt | 0.037736 | this news | 0.002136 |
| peace of mind by ensuring | 0.027778 | **a** message | 0.001539 |
| insight | 0.018868 | his answer | 0.000356 |
| the announcement | 0.017088 | **were** told | 0.000229 |
| **the** message | 0.012816 | the back and | 2.917e-05 |

- lexical variation (announcement, message, news, told, . . . )
- Morphological variation (information, informed)
- Included function words (the, a, were, this)
- Noise (the, the back and)

# Real Example

Phrase translations for **begreppet** taken from EUROPARL

| English | $\phi(\bar{t}\|\bar{s})$ | English | $\phi(\bar{t}\|\bar{s})$ |
|---|---|---|---|
| **the** | 0.226415 | the news | 0.012816 |
| told | 0.169811 | the report | 0.008544 |
| announcement | 0.075472 | the information | 0.008544 |
| message | 0.056604 | the back | 0.004272 |
| news | 0.056604 | the suspension | 0.004272 |
| information | 0.037736 | the death | 0.004272 |
| informed | 0.037736 | this announcement | 0.002848 |
| learnt | 0.037736 | this news | 0.002136 |
| peace of mind by ensuring | 0.027778 | a message | 0.001539 |
| insight | 0.018868 | his answer | 0.000356 |
| the announcement | 0.017088 | were told | 0.000229 |
| the message | 0.012816 | **the back and** | 2.917e-05 |

- lexical variation (announcement, message, news, told, . . . )
- Morphological variation (information, informed)
- Included function words (the, a, were, this)
- Noise (the, the back and)

# Goals for Today

Why phrases?

Symmetrisation of word alignment

Phrase extraction and scoring

Log-Linear Model

# Goals for Today

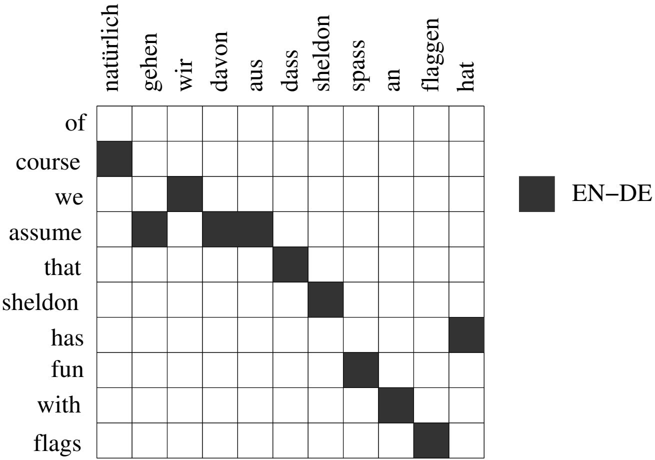
Why phrases?

**Symmetrisation of word alignment**

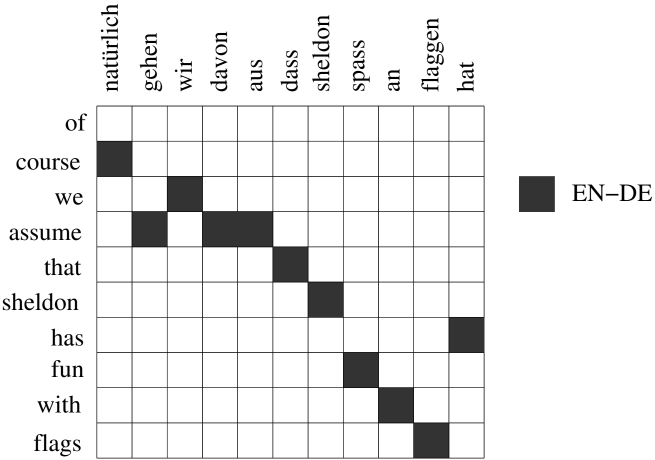
Phrase extraction and scoring

Log-Linear Model

# Directional Word Alignment Matrix

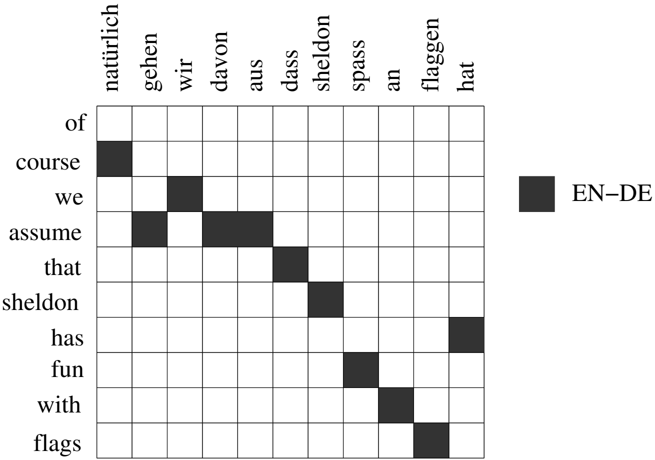


# Directional Word Alignment Matrix




Each target word can be aligned to at most one source word

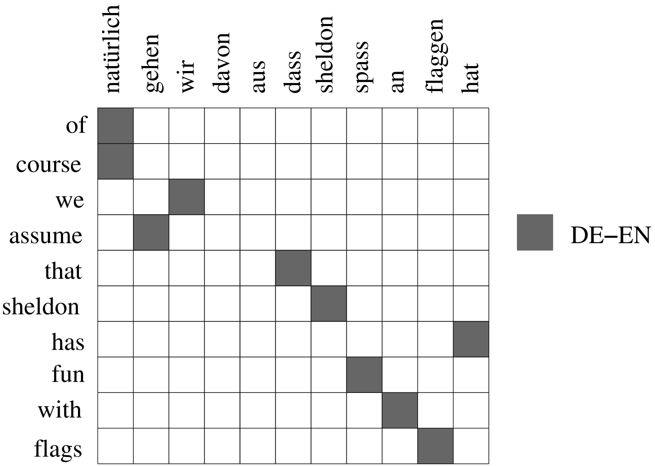
# Directional Word Alignment Matrix




Each target word can be aligned to at most one source word

**But**: a source word can be aligned to more than one target word

# Directional Word Alignment Matrix




Each target word can be aligned to at most one source word

**But**: a source word can be aligned to more than one target word

# Word Alignment Do-it-Yourself

- Languages available:
  Swedish, German, Spanish, French, Chinese
- Do **not** fill out all matrices!
- Start with the language you know best!

# SWEDISH I

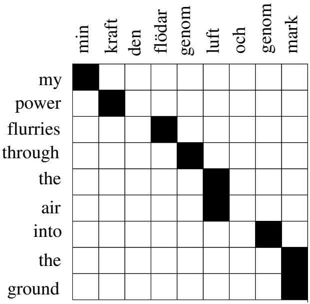
| | min | kraft | den | flödar | genom | luft | och | genom | mark |
|---|---|---|---|---|---|---|---|---|---|
| my | | | | | | | | | |
| power | | | | | | | | | |
| flurries | | | | | | | | | |
| through | | | | | | | | | |
| the | | | | | | | | | |
| air | | | | | | | | | |
| into | | | | | | | | | |
| the | | | | | | | | | |
| ground | | | | | | | | | |

Swedish $\rightarrow$ English

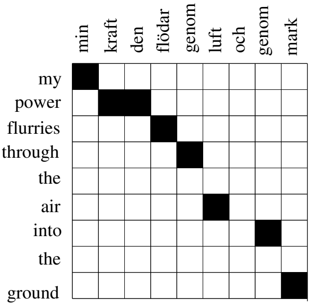
| | min | kraft | den | flödar | genom | luft | och | genom | mark |
|---|---|---|---|---|---|---|---|---|---|
| my | | | | | | | | | |
| power | | | | | | | | | |
| flurries | | | | | | | | | |
| through | | | | | | | | | |
| the | | | | | | | | | |
| air | | | | | | | | | |
| into | | | | | | | | | |
| the | | | | | | | | | |
| ground | | | | | | | | | |

English$\rightarrow$ Swedish

# SWEDISH I




Swedish → English




English→ Swedish

# SWEDISH II

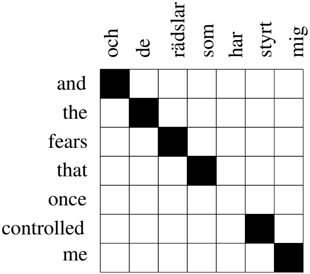
| | och | de | rädslar | som | har | styrt | mig |
|---|---|---|---|---|---|---|---|
| and | | | | | | | |
| the | | | | | | | |
| fears | | | | | | | |
| that | | | | | | | |
| once | | | | | | | |
| controlled | | | | | | | |
| me | | | | | | | |

Swedish $\rightarrow$ English

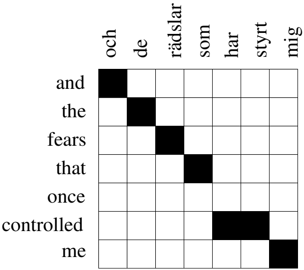
| | och | de | rädslar | som | har | styrt | mig |
|---|---|---|---|---|---|---|---|
| and | | | | | | | |
| the | | | | | | | |
| fears | | | | | | | |
| that | | | | | | | |
| once | | | | | | | |
| controlled | | | | | | | |
| me | | | | | | | |

English $\rightarrow$ Swedish

# SWEDISH II




Swedish $\rightarrow$ English




English$\rightarrow$ Swedish

# GERMAN I

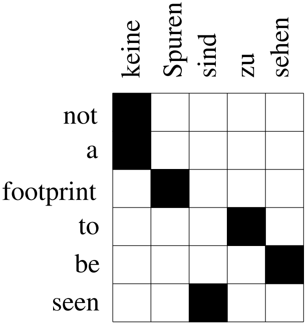
| | keine | Spuren | sind | zu | sehen |
|---|---|---|---|---|---|
| not | | | | | |
| a | | | | | |
| footprint | | | | | |
| to | | | | | |
| be | | | | | |
| seen | | | | | |

German → English

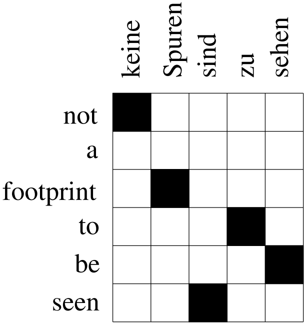
| | keine | Spuren | sind | zu | sehen |
|---|---|---|---|---|---|
| not | | | | | |
| a | | | | | |
| footprint | | | | | |
| to | | | | | |
| be | | | | | |
| seen | | | | | |

English→ German

# GERMAN I




German → English




English→ German

# GERMAN II

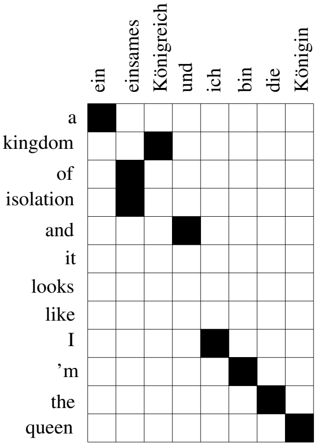
| | ein | einsames | Königreich | und | ich | bin | die | Königin |
|---|---|---|---|---|---|---|---|---|
| a | | | | | | | | |
| kingdom | | | | | | | | |
| of | | | | | | | | |
| isolation | | | | | | | | |
| and | | | | | | | | |
| it | | | | | | | | |
| looks | | | | | | | | |
| like | | | | | | | | |
| I | | | | | | | | |
| 'm | | | | | | | | |
| the | | | | | | | | |
| queen | | | | | | | | |

German $\rightarrow$ English

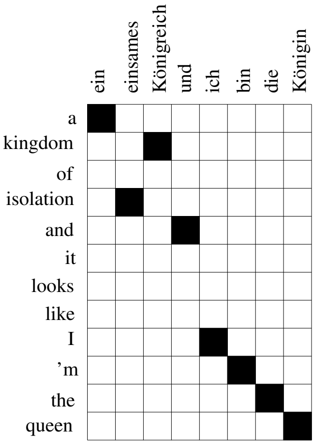
| | ein | einsames | Königreich | und | ich | bin | die | Königin |
|---|---|---|---|---|---|---|---|---|
| a | | | | | | | | |
| kingdom | | | | | | | | |
| of | | | | | | | | |
| isolation | | | | | | | | |
| and | | | | | | | | |
| it | | | | | | | | |
| looks | | | | | | | | |
| like | | | | | | | | |
| I | | | | | | | | |
| 'm | | | | | | | | |
| the | | | | | | | | |
| queen | | | | | | | | |

English$\rightarrow$ German

# GERMAN II




German → English




English → German

# SPANISH I

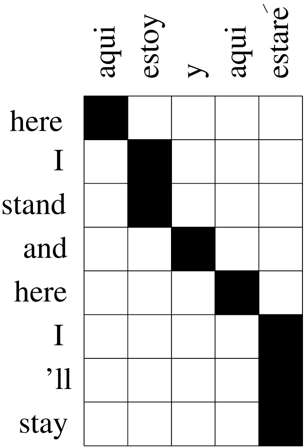
| | aqui | estoy | y | aqui | estaré |
|---|---|---|---|---|---|
| here | | | | | |
| I | | | | | |
| stand | | | | | |
| and | | | | | |
| here | | | | | |
| I | | | | | |
| 'll | | | | | |
| stay | | | | | |

Spanish → English

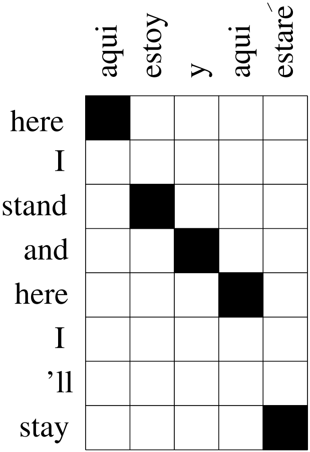
| | aqui | estoy | y | aqui | estaré |
|---|---|---|---|---|---|
| here | | | | | |
| I | | | | | |
| stand | | | | | |
| and | | | | | |
| here | | | | | |
| I | | | | | |
| 'll | | | | | |
| stay | | | | | |

English→ Spanish

# SPANISH I




Spanish → English




English→ Spanish

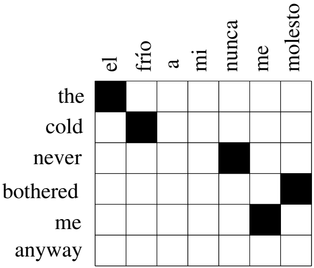
# SPANISH II

| | el | frío | a | mi | nunca | me | molesto |
|---|---|---|---|---|---|---|---|
| the | | | | | | | |
| cold | | | | | | | |
| never | | | | | | | |
| bothered | | | | | | | |
| me | | | | | | | |
| anyway | | | | | | | |

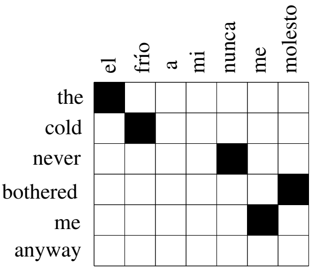
Spanish $\rightarrow$ English

| | el | frío | a | mi | nunca | me | molesto |
|---|---|---|---|---|---|---|---|
| the | | | | | | | |
| cold | | | | | | | |
| never | | | | | | | |
| bothered | | | | | | | |
| me | | | | | | | |
| anyway | | | | | | | |

English$\rightarrow$ Spanish




Spanish → English




English→ Spanish

# FRENCH

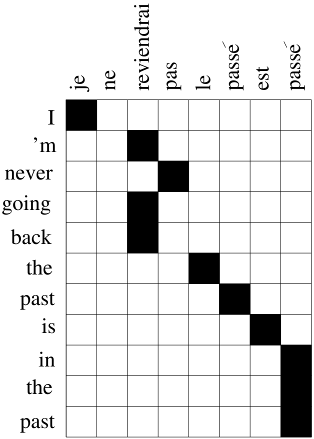
| | je | ne | reviendrai | pas | le | passé | est | passé |
|---|---|---|---|---|---|---|---|---|
| I | | | | | | | | |
| 'm | | | | | | | | |
| never | | | | | | | | |
| going | | | | | | | | |
| back | | | | | | | | |
| the | | | | | | | | |
| past | | | | | | | | |
| is | | | | | | | | |
| in | | | | | | | | |
| the | | | | | | | | |
| past | | | | | | | | |

French $\rightarrow$ English

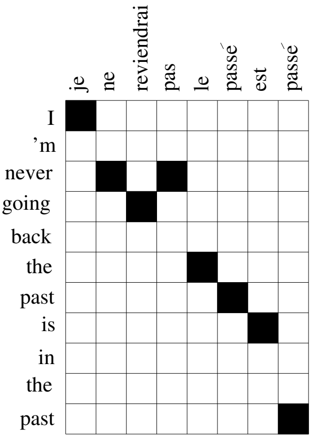
| | je | ne | reviendrai | pas | le | passé | est | passé |
|---|---|---|---|---|---|---|---|---|
| I | | | | | | | | |
| 'm | | | | | | | | |
| never | | | | | | | | |
| going | | | | | | | | |
| back | | | | | | | | |
| the | | | | | | | | |
| past | | | | | | | | |
| is | | | | | | | | |
| in | | | | | | | | |
| the | | | | | | | | |
| past | | | | | | | | |

English $\rightarrow$ French

# FRENCH




French → English




English→ French

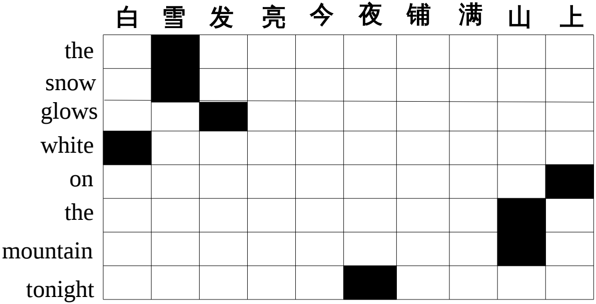
# CHINESE TO ENGLISH

| | 白 | 雪 | 发 | 亮 | 今 | 夜 | 铺 | 满 | 山 | 上 |
|---|---|---|---|---|---|---|---|---|---|---|
| the | | | | | | | | | | |
| snow | | | | | | | | | | |
| glows | | | | | | | | | | |
| white | | | | | | | | | | |
| on | | | | | | | | | | |
| the | | | | | | | | | | |
| mountain | | | | | | | | | | |
| tonight | | | | | | | | | | |

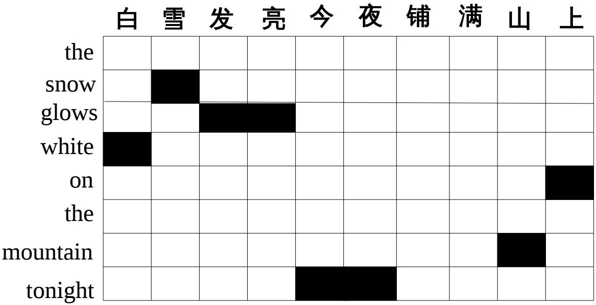
# ENGLISH TO CHINESE

| | 白 | 雪 | 发 | 亮 | 今 | 夜 | 铺 | 满 | 山 | 上 |
|---|---|---|---|---|---|---|---|---|---|---|
| the | | | | | | | | | | |
| snow | | | | | | | | | | |
| glows | | | | | | | | | | |
| white | | | | | | | | | | |
| on | | | | | | | | | | |
| the | | | | | | | | | | |
| mountain | | | | | | | | | | |
| tonight | | | | | | | | | | |

# CHINESE TO ENGLISH



# ENGLISH TO CHINESE



# Word Alignment Symmetrisation

How can we bring the two unidirectional alignments together?

- Intersection: high precision, but too few links
- Union: high recall, but too many links

- add diagonally adjacent links
- add links for unaligned words in a final step

# Word Alignment Symmetrisation

How can we bring the two unidirectional alignments together?

**Problem**:

- Intersection: high precision, but too few links
- Union: high recall, but too many links

- add diagonally adjacent links
- add links for unaligned words in a final step

# Word Alignment Symmetrisation

How can we bring the two unidirectional alignments together?

**Problem**:

- Intersection: high precision, but too few links
- Union: high recall, but too many links

- add diagonally adjacent links
- add links for unaligned words in a final step

# Word Alignment Symmetrisation

How can we bring the two unidirectional alignments together?

**Problem:**

- Intersection: high precision, but too few links
- Union: high recall, but too many links

- add diagonally adjacent links
- add links for unaligned words in a final step

# Word Alignment Symmetrisation

How can we bring the two unidirectional alignments together?

**Problem**:

- Intersection: high precision, but too few links
- Union: high recall, but too many links

**Solution**:
start from the intersection and then "**grow**"
the alignment towards the union

- add diagonally adjacent links
- add links for unaligned words in a final step

# Word Alignment Symmetrisation

How can we bring the two unidirectional alignments together?

**Problem:**

- Intersection: high precision, but too few links
- Union: high recall, but too many links

**Solution:**
start from the intersection and then "**grow**"
the alignment towards the union

- add diagonally adjacent links
- add links for unaligned words in a final step

# Word Alignment Symmetrisation

How can we bring the two unidirectional alignments together?

**Problem:**

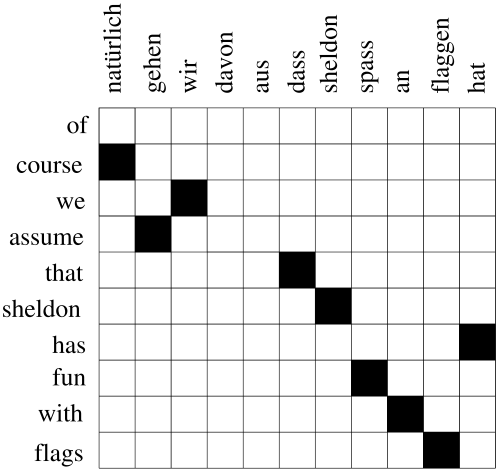
- Intersection: high precision, but too few links
- Union: high recall, but too many links

**Solution:**

start from the intersection and then "**grow**"
the alignment towards the union

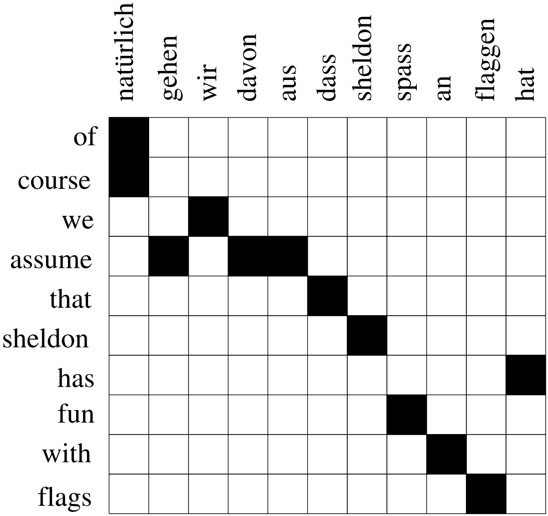
- add diagonally adjacent links
- add links for unaligned words in a final step

# Word Alignment Intersection



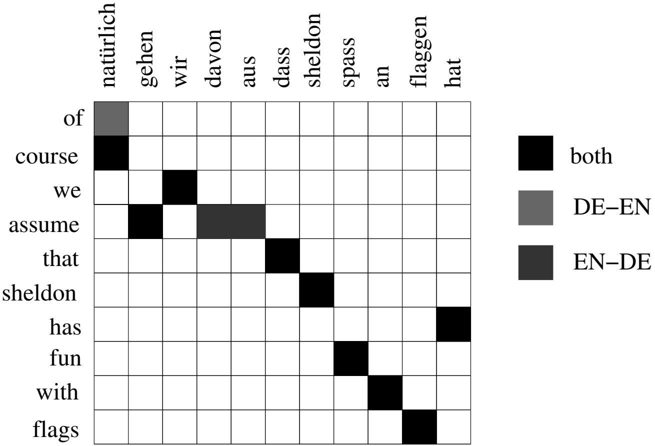

Intersection

# Word Alignment Union




Union

# Word Alignment Symmetrisation



# Goals for Today

Why phrases?

Symmetrisation of word alignment

Phrase extraction and scoring

Log-Linear Model

# Goals for Today

Why phrases?

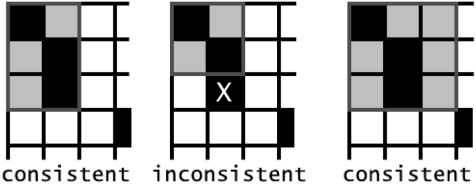
Symmetrisation of word alignment

**Phrase extraction and scoring**

Log-Linear Model

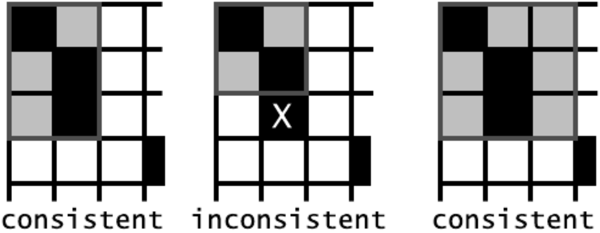
# Consistent Phrase Pairs

All words of the phrase pairs have to align to each other



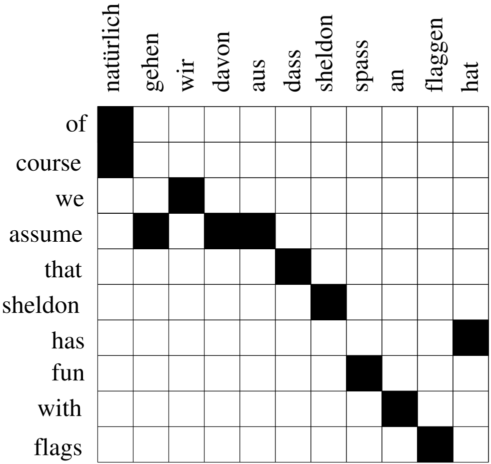

| **ok** | **violated** | **ok** |
|---|---|---|
| | one alignment point outside | unaligned word is fine |

# Phrase Extration Definition

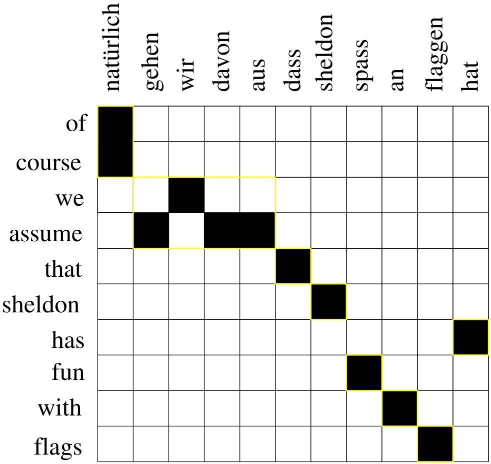



A phrase pair $(\bar{t}, \bar{s})$ is consistent with an alignment $A$,
if all words $s_1, ..., s_m$ in $\bar{s}$ that have alignment points in $A$ have these with words $t_1, ..., t_n$ in $\bar{t}$ and vice versa and at least one word in $\bar{t}$ is aligned to at least one word in $\bar{s}$.

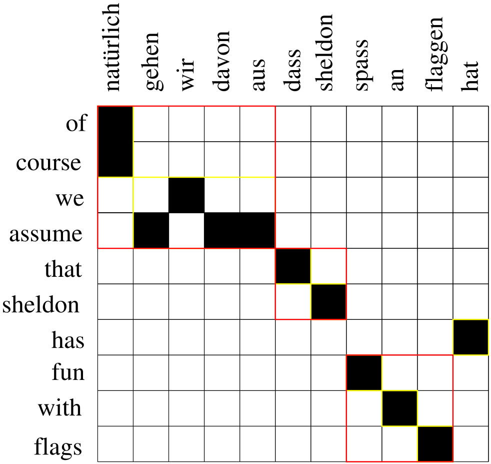
# Phrase Extraction Example



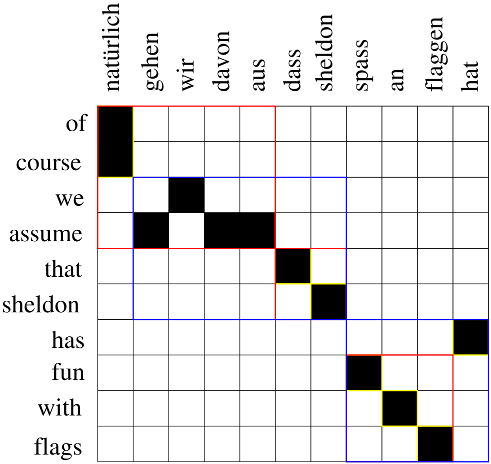
# Phrase Extraction Example



# Phrase Extraction Example



# Phrase Extraction Example



# Phrase Extraction Do-it-Yourself

Why Clickers?

- **not**: teacher controlling individual students
- **no** influence on course examination
- **instead**:
  - voluntary exercises
  - learning by doing
  - self-assessment for the students

# Phrase Extraction Do-it-Yourself

Why Clickers?

- **not**: teacher controlling individual students
- **no** influence on course examination
- **instead**:
  - voluntary exercises
  - learning by doing
  - self-assessment for the students

# Phrase Extraction Do-it-Yourself

Why Clickers?

- **not**: teacher controlling individual students
- **no** influence on course examination
- **instead**:
  - voluntary exercises
  - learning by doing
  - self-assessment for the students

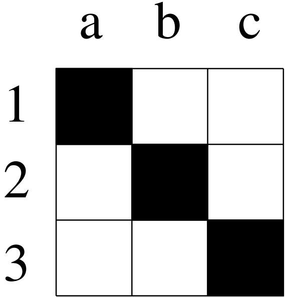
# Phrase Extraction Do-it-Yourself

Why Clickers?

- **not**: teacher controlling individual students
- **no** influence on course examination
- **instead**:
  - voluntary exercises
  - learning by doing
  - self-assessment for the students

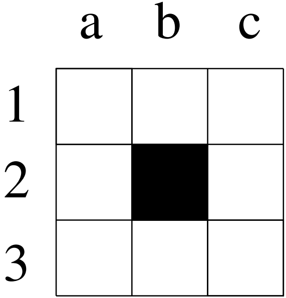
# Phrase Extraction Exercise 1

How many phrases can be extracted from this word alignment?



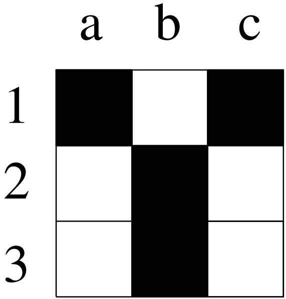
# Phrase Extraction Exercise 2

How many phrases can be extracted from this word alignment?



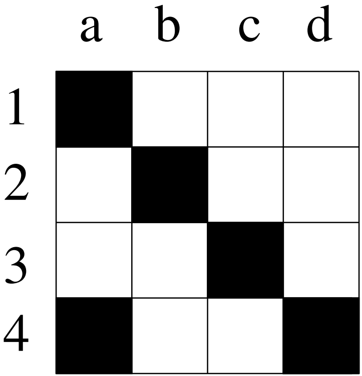
# Phrase Extraction Exercise 3

How many phrases can be extracted from this word alignment?



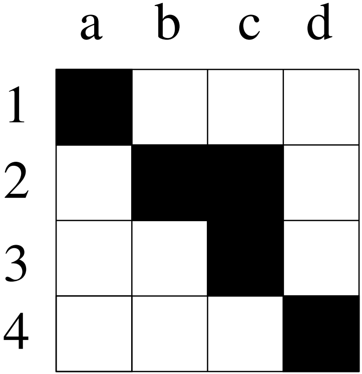
# Phrase Extraction Exercise 4

How many phrases can be extracted from this word alignment?



# Phrase Extraction Exercise 5

How many phrases can be extracted from this word alignment?



# Scoring Phrase Translations

- Phrase pair extration: collect all phrase pairs from the data
- Phrase pair scoring: assign probabilities to phrase translations
- Score by relative frequency (MLE):
  $\phi(\bar{t}|\bar{s}) = \frac{count(\bar{s},\bar{t})}{\sum_{\bar{t}_i} count(\bar{s},\bar{t})}$
- Potentially improve scoring by smoothing

# Scoring Phrase Translations

- Phrase pair extration: collect all phrase pairs from the data
- Phrase pair scoring: assign probabilities to phrase translations
- Score by relative frequency (MLE):
  $\phi(\bar{t}|\bar{s}) = \frac{count(\bar{s},\bar{t})}{\sum_{\bar{t}_i} count(\bar{s},\bar{t})}$
- Potentially improve scoring by smoothing

# Scoring Phrase Translations

- Phrase pair extration: collect all phrase pairs from the data
- Phrase pair scoring: assign probabilities to phrase translations
- Score by relative frequency (MLE):
  $\phi(\bar{t}|\bar{s}) = \frac{count(\bar{s},\bar{t})}{\sum_{\bar{t}_i} count(\bar{s},\bar{t})}$
- Potentially improve scoring by smoothing

# Scoring Phrase Translations

- Phrase pair extration: collect all phrase pairs from the data
- Phrase pair scoring: assign probabilities to phrase translations
- Score by relative frequency (MLE):
  $\phi(\bar{t}|\bar{s}) = \frac{count(\bar{s},\bar{t})}{\sum_{\bar{t}_i} count(\bar{s},\bar{t})}$
- Potentially improve scoring by smoothing

# Scoring Phrase Translations

- Phrase pair extration: collect all phrase pairs from the data
- Phrase pair scoring: assign probabilities to phrase translations
- Score by relative frequency (MLE):
  $\phi(\bar{t}|\bar{s}) = \frac{count(\bar{s},\bar{t})}{\sum_{\bar{t}_i} count(\bar{s},\bar{t})}$
- Potentially improve scoring by smoothing

# Goals for Today

Why phrases?

Symmetrisation of word alignment

Phrase extraction and scoring

Log-Linear Model

# Goals for Today

Why phrases?

Symmetrisation of word alignment

Phrase extraction and scoring

**Log-Linear Model**

# Weighted models

- The standard model consists of three sub-models:
  - Phrase translation model $\phi(\bar{s}|\bar{t})$
  - Reordering model $d$
  - Language model $p_{LM}(t)$

$$t_{best} = \arg\max_t \prod_{i=1}^{I} \phi(\bar{s}_i|\bar{t}_i) d(\text{start}_i - \text{end}_{i-1} - 1) \prod_{i=1}^{|t|} p_{LM}(t_i|t_{i-(n-1)} \dots t_{i-1})$$

- Some sub-models may be more important than others
- Add weights $\lambda_\phi, \lambda_d, \lambda_{LM}$

# Weighted models

- The standard model consists of three sub-models:
  - Phrase translation model $\phi(\bar{s}|\bar{t})$
  - Reordering model $d$
  - Language model $p_{LM}(t)$

$$t_{best} = \arg\max_t \prod_{i=1}^{I} \phi(\bar{s}_i|\bar{t}_i) d(\text{start}_i - \text{end}_{i-1} - 1) \prod_{i=1}^{|t|} p_{LM}(t_i|t_{i-(n-1)} \dots t_{i-1})$$

- Some sub-models may be more important than others
- Add weights $\lambda_\phi, \lambda_d, \lambda_{LM}$

# Weighted models

- The standard model consists of three sub-models:
  - Phrase translation model $\phi(\bar{s}|\bar{t})$
  - Reordering model $d$
  - Language model $p_{LM}(t)$

$$t_{best} = \arg\max_t \prod_{i=1}^{I} \phi(\bar{s}_i|\bar{t}_i) d(\text{start}_i - \text{end}_{i-1} - 1) \prod_{i=1}^{|t|} p_{LM}(t_i|t_{i-(n-1)} \dots t_{i-1})$$

- Some sub-models may be more important than others
- Add weights $\lambda_\phi, \lambda_d, \lambda_{LM}$

# Weighted models

- The standard model consists of three sub-models:
  - Phrase translation model $\phi(\bar{s}|\bar{t})$
  - Reordering model $d$
  - Language model $p_{LM}(t)$

$$t_{best} = \arg\max_t \prod_{i=1}^{I} \phi(\bar{s}_i|\bar{t}_i)^{\lambda_\phi} d(\text{start}_i - \text{end}_{i-1} - 1)^{\lambda_d} \prod_{i=1}^{|t|} p_{LM}(t_i|t_{i-(n-1)} \dots t_{i-1})^{\lambda_{LM}}$$

- Some sub-models may be more important than others
- Add weights $\lambda_\phi, \lambda_d, \lambda_{LM}$

# Log-linear models

- Such a weighted model is a log-linear model:

$$p(x) = \exp \sum_{i=1}^{n} \lambda_i h_i(x)$$

- Our feature functions:
  - three feature functions $n = 3$
  - random variable $x = (s, t, \text{start}, \text{end})$
  - feature function $h_1 = \log \phi$
  - feature function $h_2 = \log d$
  - feature function $h_3 = \log p_{LM}$

# Weighted model as a log-linear model

$$
\begin{aligned}
p(t, a|s) = \exp(\lambda_\phi \sum_{i=1}^{I} & \log \phi(\bar{s}_i|\bar{t}_i) + \\
\lambda_d \sum_{i=1}^{I} & \log d(\text{start}_i - \text{end}_{i-1} - 1) + \\
\lambda_{LM} \sum_{i=1}^{|t|} & \log p_{LM}(t_i | t_{i-(n-1)} \ldots t_{i-1}))
\end{aligned}
$$

# More feature functions

$$t^* = \arg\max_t \sum_i \lambda_i h_i(s, t)$$

- Easy and useful to add more feature functions
  - Bidirectional alignment probabilities $\phi(\bar{s}|\bar{t})$ and $\phi(\bar{t}|\bar{s})$
  - Lexical weighting of phrase pairs:
    useful since rare phrase pairs have unreliable probability estimates

# More feature functions

$$t^* = \arg\max_t \sum_i \lambda_i h_i(s, t)$$

- Easy and useful to add more feature functions
  - Bidirectional alignment probabilities $\phi(\bar{s}|\bar{t})$ and $\phi(\bar{t}|\bar{s})$
  - Lexical weighting of phrase pairs:
    useful since rare phrase pairs have unreliable probability estimates

# More feature functions

$$t^* = \arg\max_t \sum_i \lambda_i h_i(s, t)$$

- Easy and useful to add more feature functions
  - Bidirectional alignment probabilities $\phi(\bar{s}|\bar{t})$ and $\phi(\bar{t}|\bar{s})$
  - Lexical weighting of phrase pairs:
    useful since rare phrase pairs have unreliable probability estimates

# More feature functions 2

- Language model has a bias towards short translations
  - word count: $\text{wc}(t) = \log |t|^{\omega}$
- We may prefer finer or coarser segmentations
  - phrase count: $\text{pc}(t) = \log |I|^{\rho}$
- Multiple language models
- Other knowledge sources

# More feature functions 2

- Language model has a bias towards short translations
  - word count: $\mathrm{wc}(t) = \log |t|^{\omega}$
- We may prefer finer or coarser segmentations
  - phrase count: $\mathrm{pc}(t) = \log |I|^{\rho}$
- Multiple language models
- Other knowledge sources

# More feature functions 2

- Language model has a bias towards short translations
  - word count: $\mathrm{wc}(t) = \log |t|^{\omega}$
- We may prefer finer or coarser segmentations
  - phrase count: $\mathrm{pc}(t) = \log |I|^{\rho}$
- Multiple language models
- Other knowledge sources

# More feature functions 2

- Language model has a bias towards short translations
  - word count: $\text{wc}(t) = \log |t|^{\omega}$
- We may prefer finer or coarser segmentations
  - phrase count: $\text{pc}(t) = \log |I|^{\rho}$
- Multiple language models
- Other knowledge sources

# Mid-Course Feedback

Sara asked me to ask you for feedback about the MT course so far

What did you like? What should be improved?

We will try to react to your feedback during the rest of the course!

# Information for Bachelor's Students

Assignment 1 is now online.

You have 2 weeks from now to hand in this assignment.

Contact Sara if you have any further questions!