Machine Translation
Reordering

Miles Osborne
(based on slides by Barry Haddow and Philipp Koehn)

Informatics

Feb 2013

School of informatics

Reordering

- Different languages put words in different order
- Computational Problem
  - during translation we have to compose an output sentence
  - when composing a sentence of n words, there are $n!$ possible orderings
  - exponential spaces are hard to search
- Linguistic Problem
  - how do different languages define the word order?
  - how can we model reordering transformations during translation?

School of informatics

Reordering

- Reordering remains a hard problem for SMT and motives new models
- Reordering is the most important feature for predicting translation performance (on indo-european languages)

School of informatics

Reordering performance

- Reordering performance examined in [Birch et al, 2009]
- Compared phrase-based and hierarchical
  - hiero rules e.g. ich hat X gegessen → I have eaten X
- Hierarchical better at medium range
- Phrase-based better at short range
- Neither did well at long range

School of informatics

Bag of Word Problem

the nobody security Swiss account the of will bank trust

- From New York Times
- Can you put the words in the correct order?
  - → Does a good language model suffice?

Bag of Word Problem

nobody will trust the security of the Swiss bank account

- From New York Times
- Can you put the words in the correct order?
  - → Does a good language model suffice?

School of informatics
Miles Osborne (Informatics) Machine Translation Feb 2013 5 / 27

The Language Model/Translation Model Debate

- Language model argument
  - Humans can make sense of scrambled sentences, machines should too
  - Build really big language models to determine what makes sense in the target language
- Translation model argument
  - structure in source sentence defines relationship between words
  - we can learn how to map this structure into a target-language structure

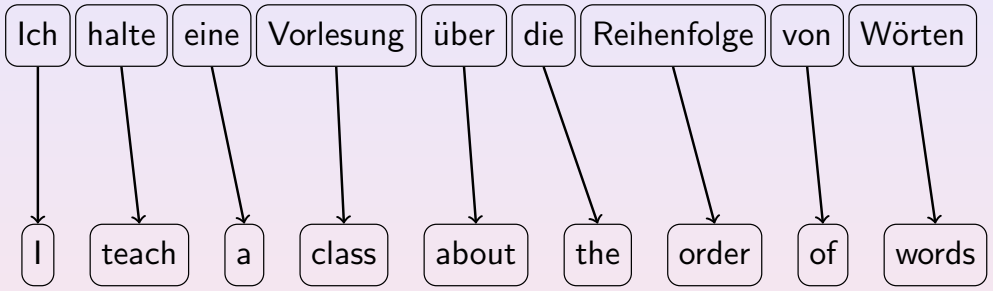
School of informatics

The Language Model/Translation Model Debate

- LM reordering is not directly connected to the source
  - . . . but we have a lot of evidence for it
- TM reordering is connected with the source
  - . . . but we have a less evidence for it

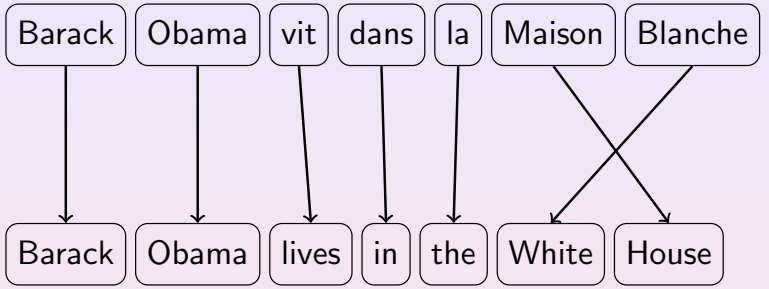
School of informatics

Examples I



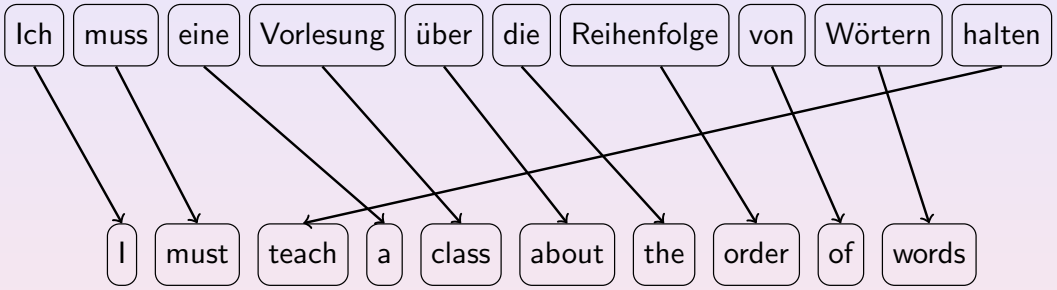
Monotone reordering

Examples II



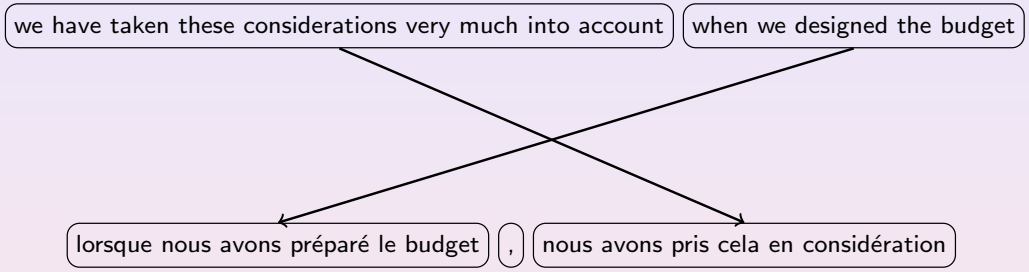
French-English adjective/noun swapping

Examples III



German main verb movement

Examples IV



Clause reordering

Free word-order languages

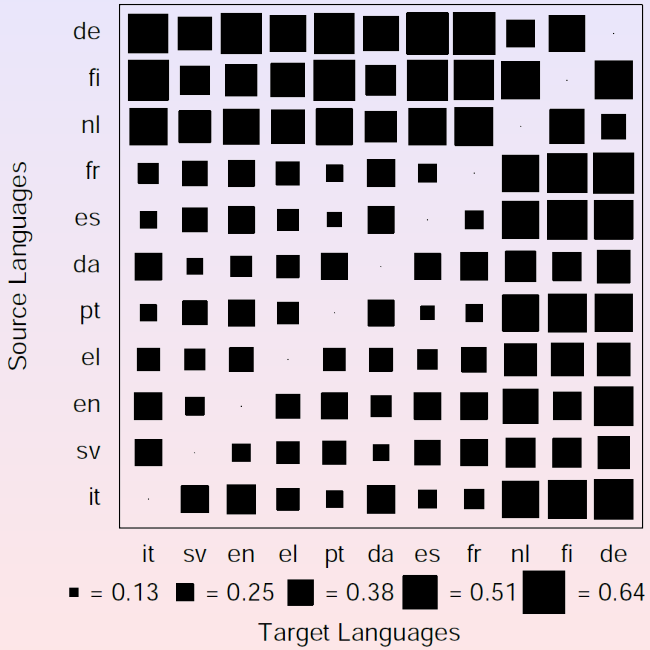
- The following German sentences mean the same:

  Der Mann gibt der Frau das Buch.
  Das Buch gibt der Mann der Frau.
  Der Frau gibt der Mann das Buch.
  Der Mann gibt das Buch der Frau.
  Das Buch gibt der Frau der Mann.
  Der Frau gibt das Buch der Mann.

- Are free word order languages really free?

School of informatics

Amount of reordering (European languages)



RQuantity measures the amount of reordering

Very variable across languages

[from Birch et al., EMNLP 2008]

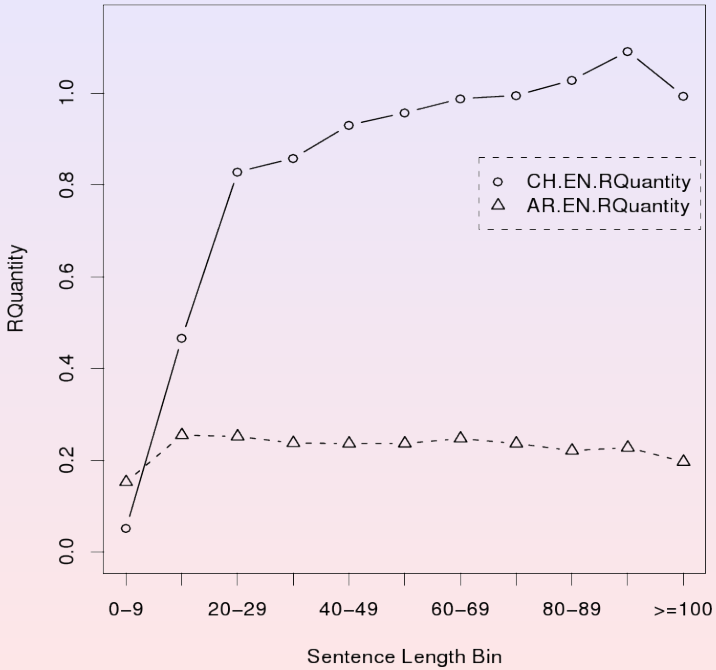
School of informatics

Amount of reordering (European languages)

Italian	it
Swedish	sv
English	en
Greek	el
Portuguese	pt
Danish	da
Spanish	es
French	fr
*Dutch	nl
*Finnish	fi
*German	de

* hard languages for reordering

Amount of reordering (Chinese vs Arabic)



Reordering against sentence length

Chinese has very different structure to English

[from Birch et al., WMT 2009]

Reordering Models in Phrase-based MT

Implicit

Language model and phrase internal

Distance

Linear function of the distance moved

Lexicalised

Each phrase pair has a probability distribution over jump types

Pre-reordering

Preprocess the source to make it more like the target language

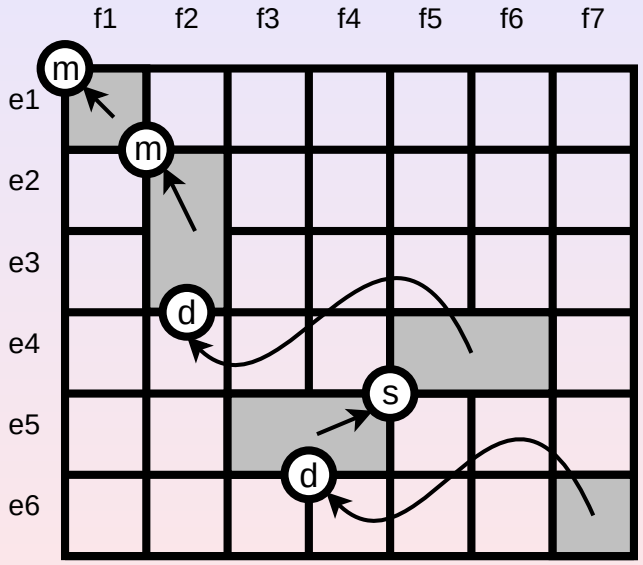
School of informatics

Distance-based Reordering

- Distance-based reordering models have a simplistic view of language
  - Consider how many words a phrase moves (eg a maximum of 6 words)
- In reality, movement depends on the words
  - German modal verbs
  - French adjectives
- Use the aligned training corpus to estimate a model

School of informatics

Lexicalised reordering models



[from Koehn et al., 2005, IWSLT]

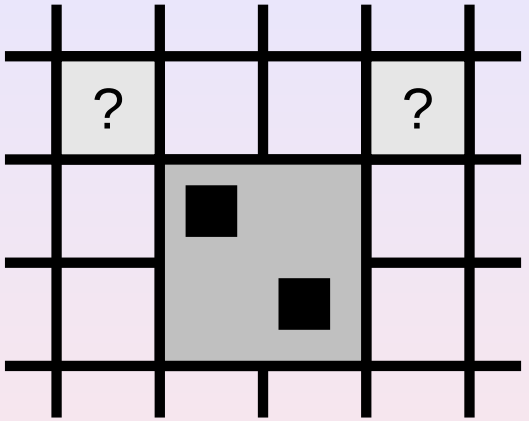
Three types: **monotone**, **swap**, **discontinuous**

Condition on phrase pair : $p(o|e, f)$

Gives six features (3 orientations, current and next phrase pair).

School of informatics

Learning lexicalised reordering models



[from Koehn et al., 2005, IWSLT]

- Collect orientation statistics during *phrase extraction*
  - Alignment *top left* ⇒ *monotone*
  - Alignment *top right* ⇒ *swap*
  - *Neither* ⇒ *discontinuous*

School of informatics

Discriminative training of lexicalised reordering

- Lexicalised reordering uses a very simple estimation strategy

$$p(o|e,f) = \frac{c(o,e,f)}{\sum_{o'} c(o',e,f)}$$

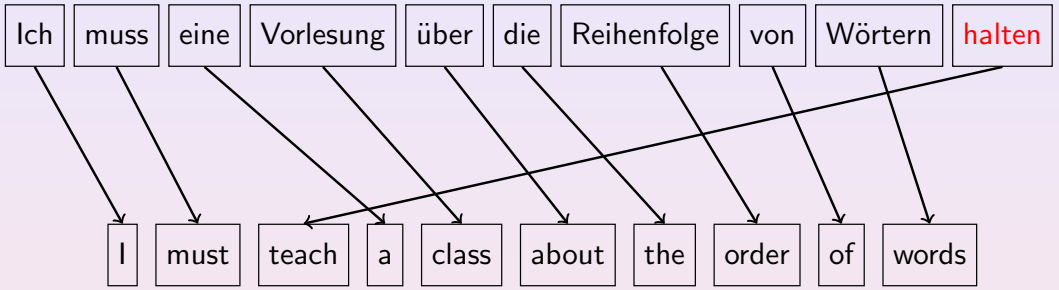
- Instead we could express the probability using a log-linear model

$$p(o|e,f) \propto \exp\left(\sum_{i=1}^{m} \mu_i \cdot g_i(e,f)\right)$$

- This makes it possible to add many knowledge sources
  - e.g. part-of-speech tags
- [Zens and Ney, WMT 2006]

School of informatics

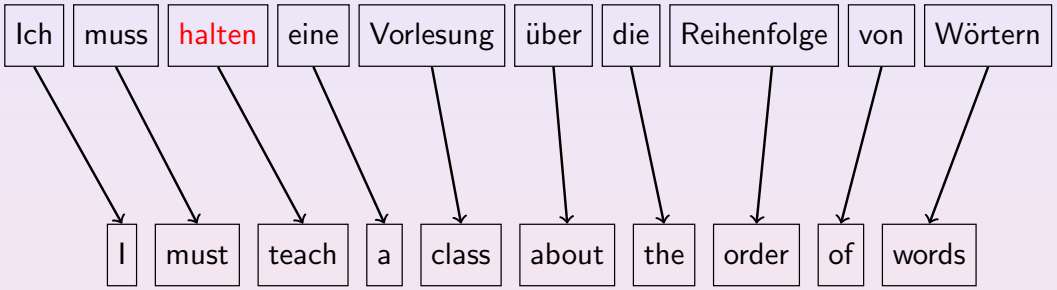
Pre-reordering: Idea



Large movement between German and English

School of informatics

Pre-reordering: Idea



Rewrite German to make it more like English

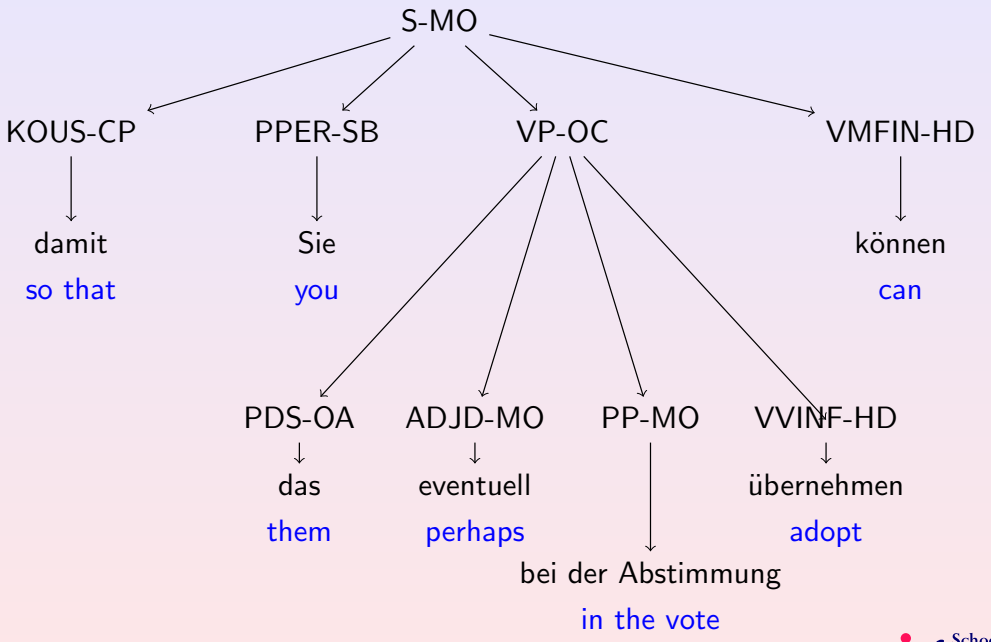
School of informatics

Clause restructuring

- Transformations are applied to all source texts
  - Training, tuning, testing, ...
- Based on parse of source text
- Original work used hand-written rules
  - [Collins et al., ACL 2005]
- Later authors tried to learn from data

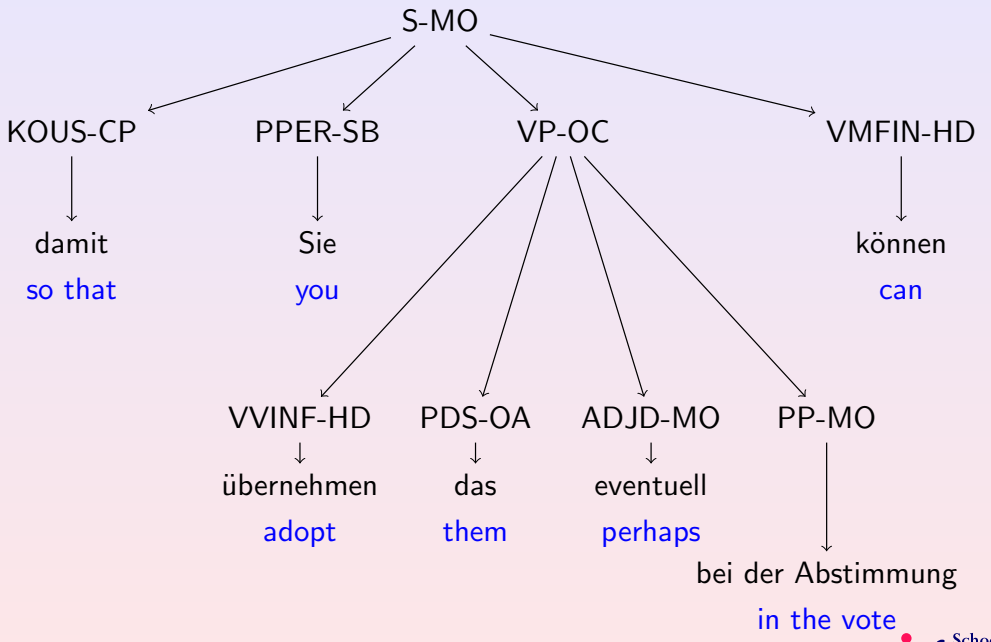
School of informatics

Clause restructuring: Example



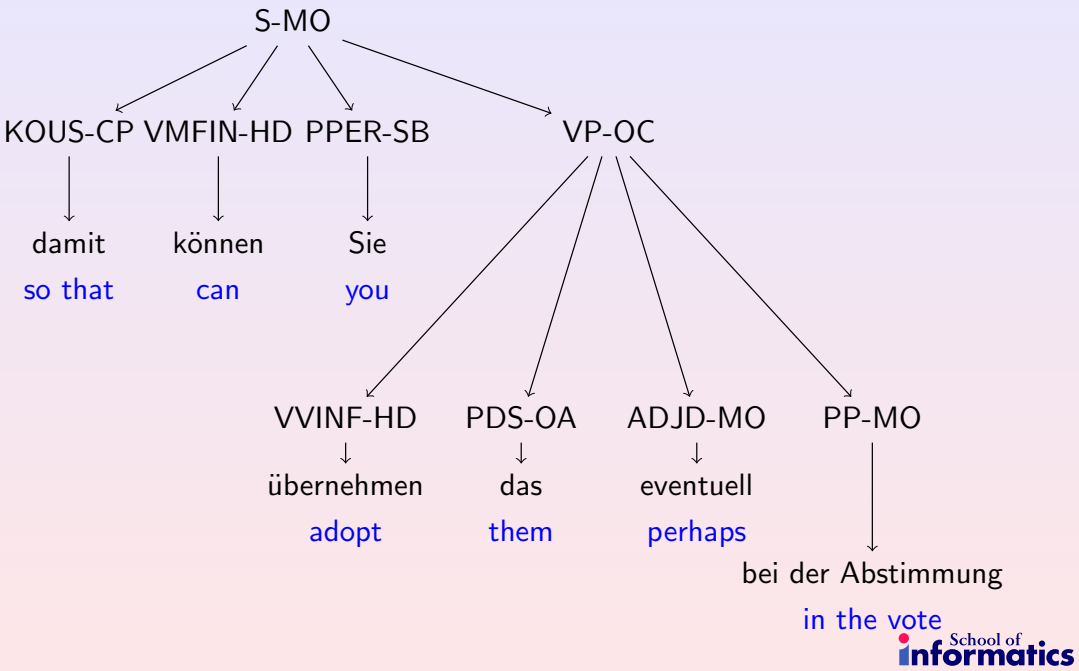
School of informatics

Clause restructuring: Verb initial

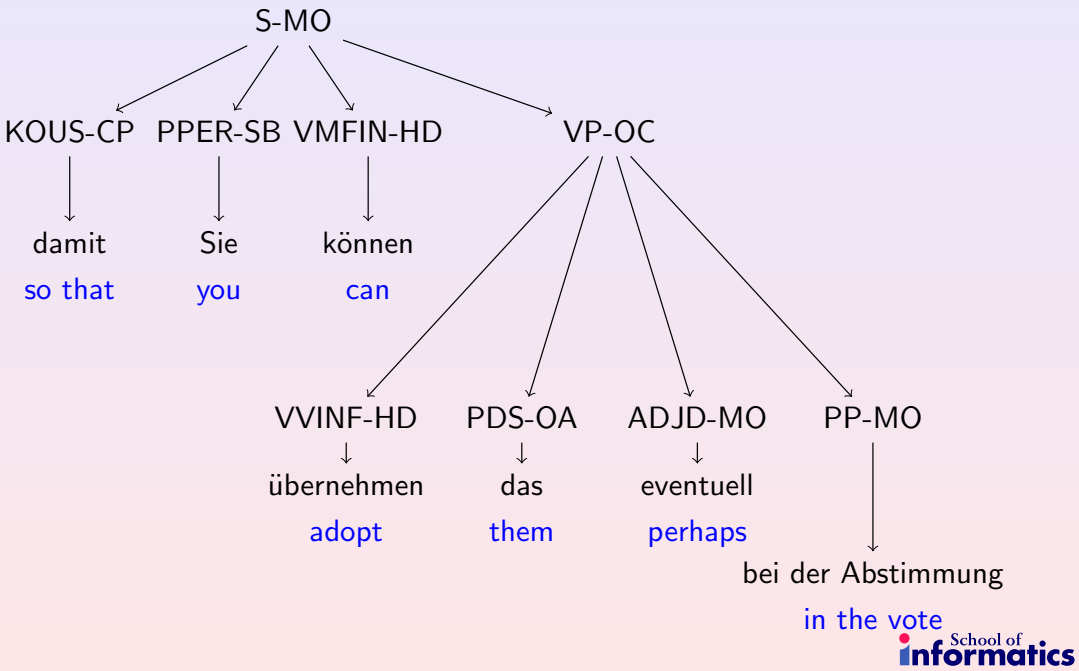


School of informatics

Clause restructuring: Verb 2nd



Clause restructuring: Move subject



Summary

- Reordering in translation
- Models for reordering in phrase-based MT
  - Distance-based
  - Lexicalised
  - Pre-reordering

School of informatics