

---

# MOSES

## Machine Translation with Open Source Software

Philipp Koehn and Hieu Hoang

4 November 2010



# Outline



**09:30-10:00 Introduction**

**10:00-11:00 Hands-on Session** — you will need a laptop

**11:00-11:30 Break**

**11:30-12:30 Advanced Topics**

# Basic Idea



# Statistical Machine Translation History



## **around 1990**

Pioneering work at IBM, inspired by success in speech recognition

## **1990s**

Dominance of IBM's word-based models, support technologies

## **early 2000s**

Phrase-based models

## **late 2000s**

Tree-based models

# Moses History

**2002**

Pharaoh decoder, precursor to Moses (phrase-based models)

**2005**

Moses started by Hieu Hoang and Philipp Koehn (factored models)

**2006**

JHU workshop extends Moses significantly

**since late 2006**

Funding by EU projects EuroMatrix, EuroMatrixPlus

**2009**

Tree-based models implemented in Moses

# Moses in Academia

- Built by academics, for academics
- Reference implementation of state of the art
  - researchers develop new methods on top of Moses
  - developers re-implement published methods
  - used by other researchers as black box
- Baseline to beat
  - researchers compare their method against Moses

# Developer Community

- Main development at University of Edinburgh, but also:
  - Fondazione Bruno Kessler (Italy)
  - Charles University (Czech Republic)
  - DFKI (Germany)
  - RWTH Aachen (Germany)
  - others...
- Code shared on Sourceforge
- Main forum: support and developer mailing lists
- Main event: Machine Translation Marathon (next: September 2011, Trento)
  - annual open source convention
  - presentation of new open source tools
  - hands-on work on new open source projects
  - summer school for statistical machine translation

# Open Source Components



- Moses distribution uses external open source tools
  - word alignment: GIZA++, Berkeley aligner
  - language model: SRILM, IRSTLM, RANDLM
  - scoring: BLEU, TER, METEOR
- Other useful tools
  - sentence aligner
  - syntactic parsers
  - part-of-speech taggers
  - morphological analyzers



## Other Open Source MT Systems

- **Joshua** — Johns Hopkins University  
<http://joshua.sourceforge.net/>
- **CDec** — University of Maryland  
<http://cdec-decoder.org/>
- **Jane** — RWTH Aachen  
<http://www-i6.informatik.rwth-aachen.de/jane/>
- Very similar technology
  - Joshua implemented in Java, others in C++
  - Joshua and Jane support only tree-based models
  - Phrasal supports only phrase-based models
- Open sourcing tools increasing trend in NLP research

# Moses in Industry

- Distributed with LGPL — free to use
- Competitive with commercial SMT solutions  
(Language Weaver, Google, ...)
- But:
  - not easy to use
  - requires significant expertise for optimal performance
  - integration into existing workflow not straight-forward

# Case Studies

## **European Commission —**

uses Moses in-house to aid human translators

## **Autodesk —**

showed productivity increases in translating manuals when post-editing output from a custom-build Moses system

## **Systran —**

developed statistical post-editing using Moses

## **Asia Online —**

offers translation technology and services based on Moses

## **Pangea —**

language service provider builds Moses systems for its data

## Phrase-Based Model



- Foreign input is segmented in phrases
- Each phrase is translated into English
- Phrases are reordered

# Phrase Translation Options

er	geht	ja	nicht	nach	hause
he	is	yes	not	after	house
it	are	is	do not	to	home
, it	goes	, of course	does not	according to	chamber
, he	go	,	is not	in	at home
it is		not		home	
he will be		is not		under house	
it goes		does not		return home	
he goes		do not		do not	
	is		to		
	are		following		
	is after all		not after		
	does		not to		
	not				
	is not				
	are not				
	is not a				

- Many translation options to choose from

# Phrase Translation Options

er	geht	ja	nicht	nach	hause
he	is	yes	not	after	house
it	are	is	do not	to	home
, it	goes	, of course	does not	according to	chamber
, he	go		is not	in	at home
it is		not		home	
he will be		is not		under house	
it goes		does not		return home	
he goes		do not		do not	
	is		to		
	are		following		
	is after all		not after		
	does		not to		
	not				
	is not				
	are not				
	is not a				

- The machine translation decoder does not know the right answer
    - picking the right translation options
    - arranging them in the right order
- Search problem solved by heuristic beam search

# Decoding: Precompute Translation Options<sup>14</sup>



consult phrase translation table for all input phrases

# Decoding: Start with Initial Hypothesis

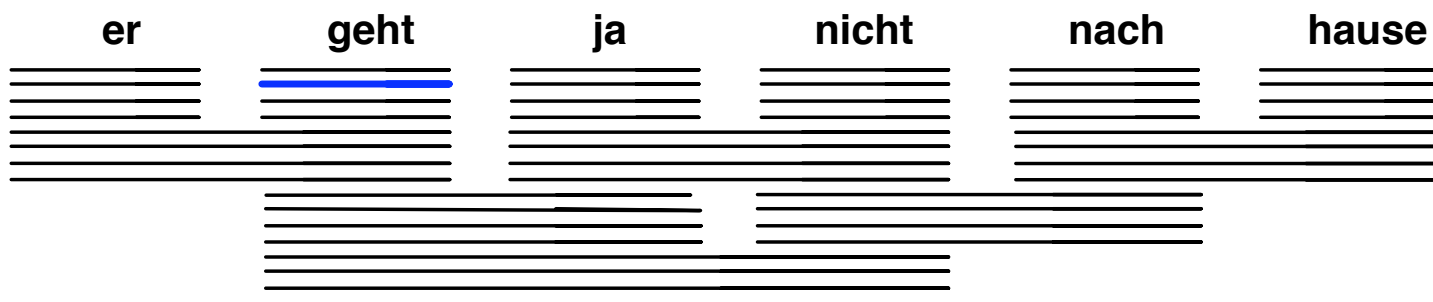
15



initial hypothesis: no input words covered, no output produced

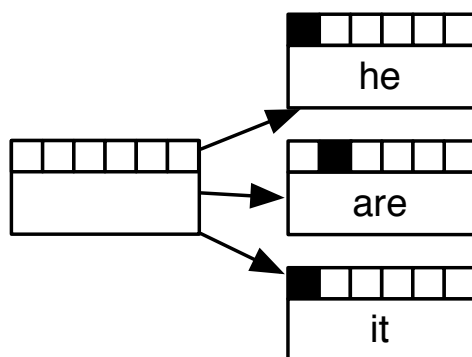
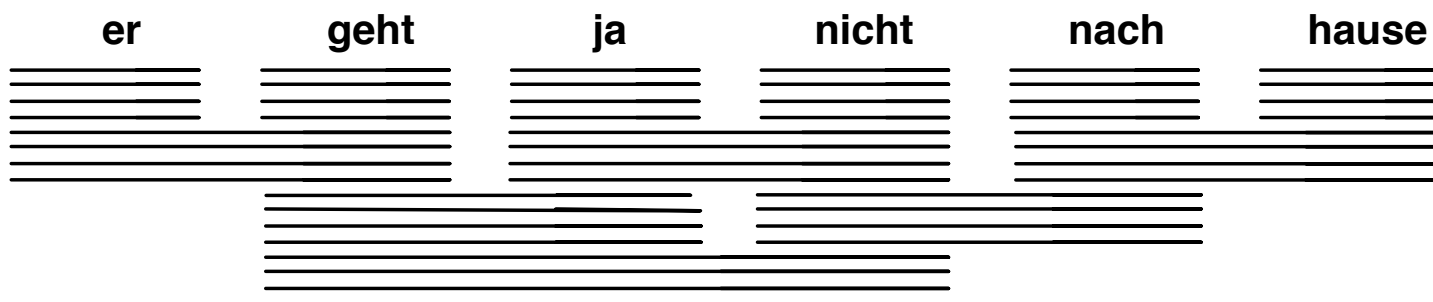


# Decoding: Hypothesis Expansion



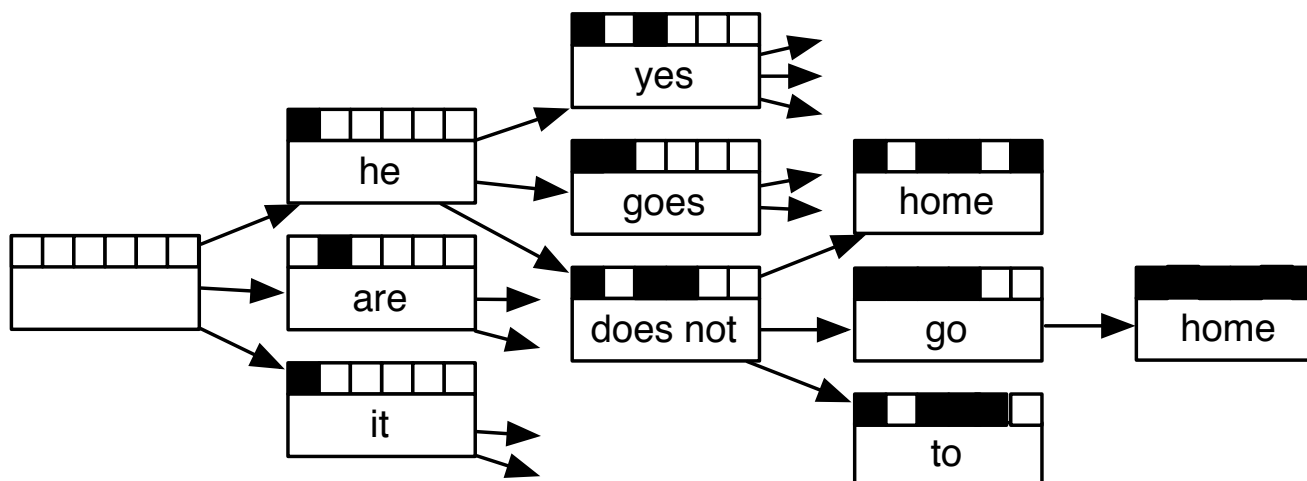
pick any translation option, create new hypothesis

# Decoding: Hypothesis Expansion



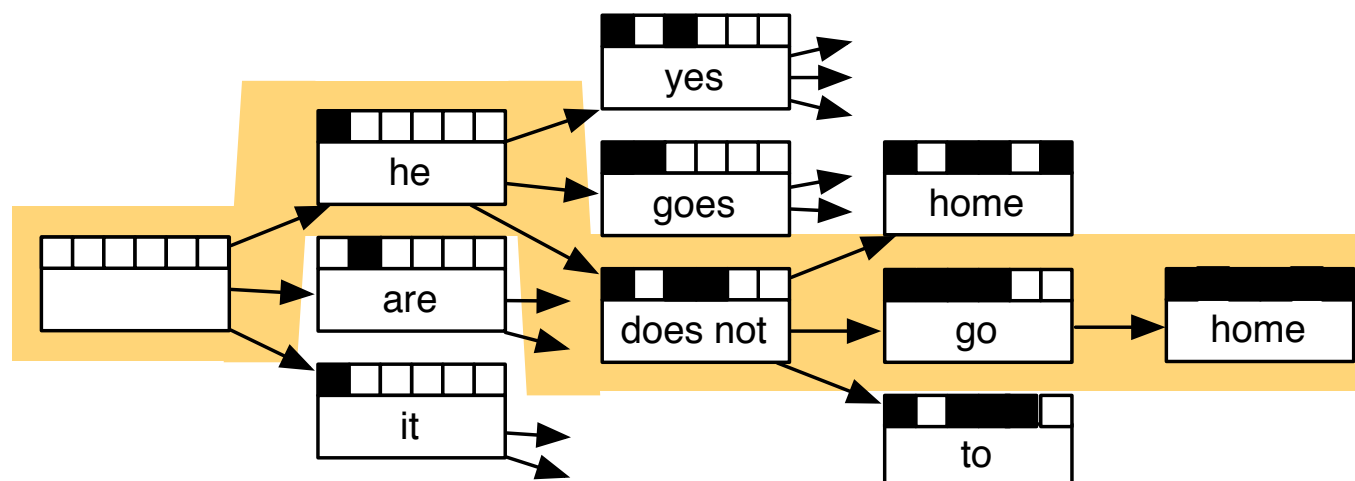
create hypotheses for all other translation options

# Decoding: Hypothesis Expansion



also create hypotheses from created partial hypothesis

# Decoding: Find Best Path



backtrack from highest scoring complete hypothesis

# Computational Complexity



- The suggested process creates exponential number of hypothesis
  - Reduction of search space: pruning
- Decoder may not find the model-best translation

# Factored Representation

- Factored representation of words



- Goals
  - generalization, e.g. by translating lemmas, not surface forms
  - richer model, e.g. using syntax for reordering, language modeling)

# Factored Model

Example:



Decomposing the translation step

Translating lemma and morphological information more robust

# Syntax Models

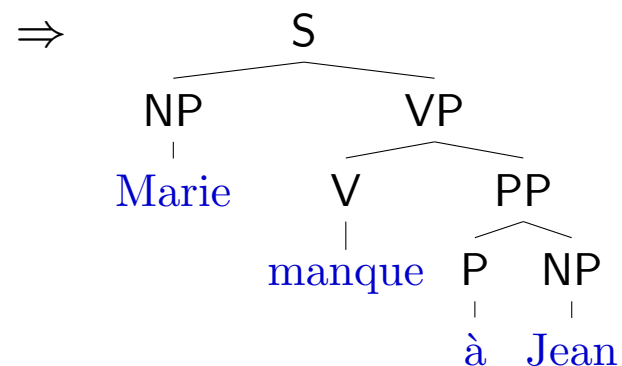
## String to String

John misses Mary

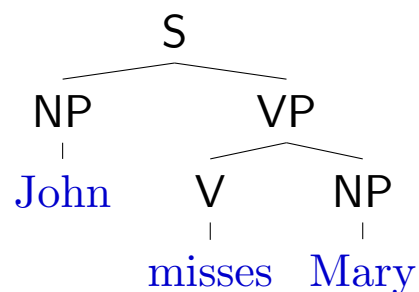
⇒ Marie manque à Jean

## String to Tree

John misses Mary

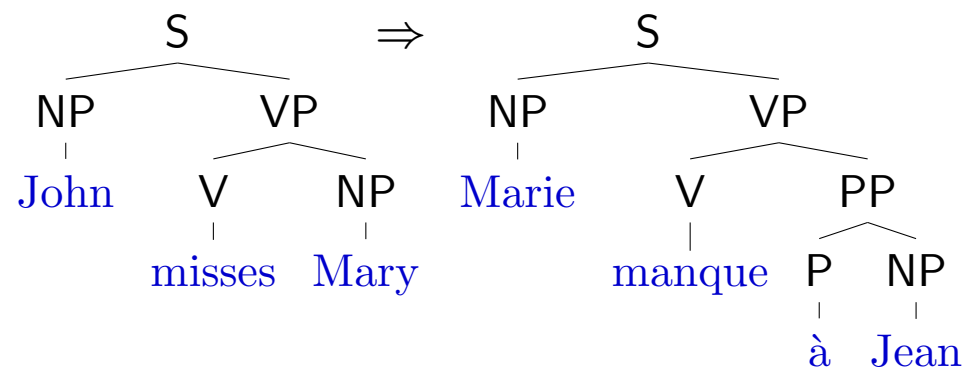


## Tree to String



⇒ Marie manque à Jean

## Tree to Tree





# Syntax Decoding



# Syntax Decoding



# Syntax Decoding

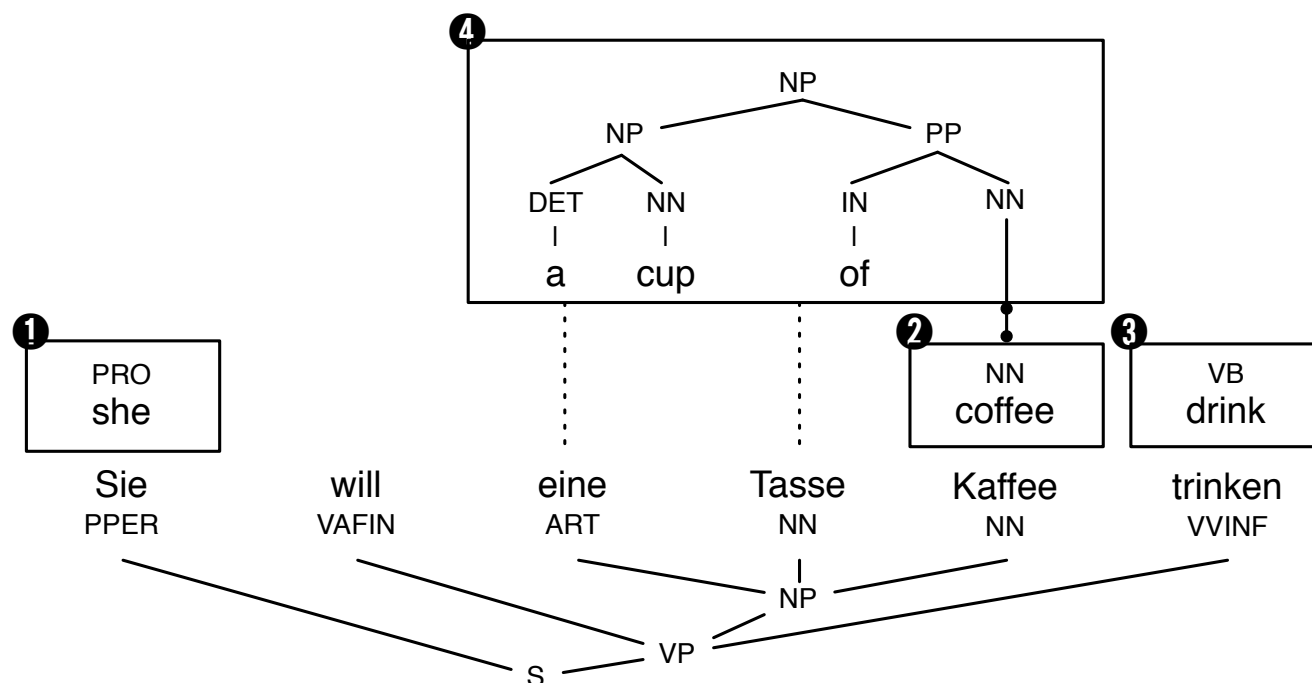
26



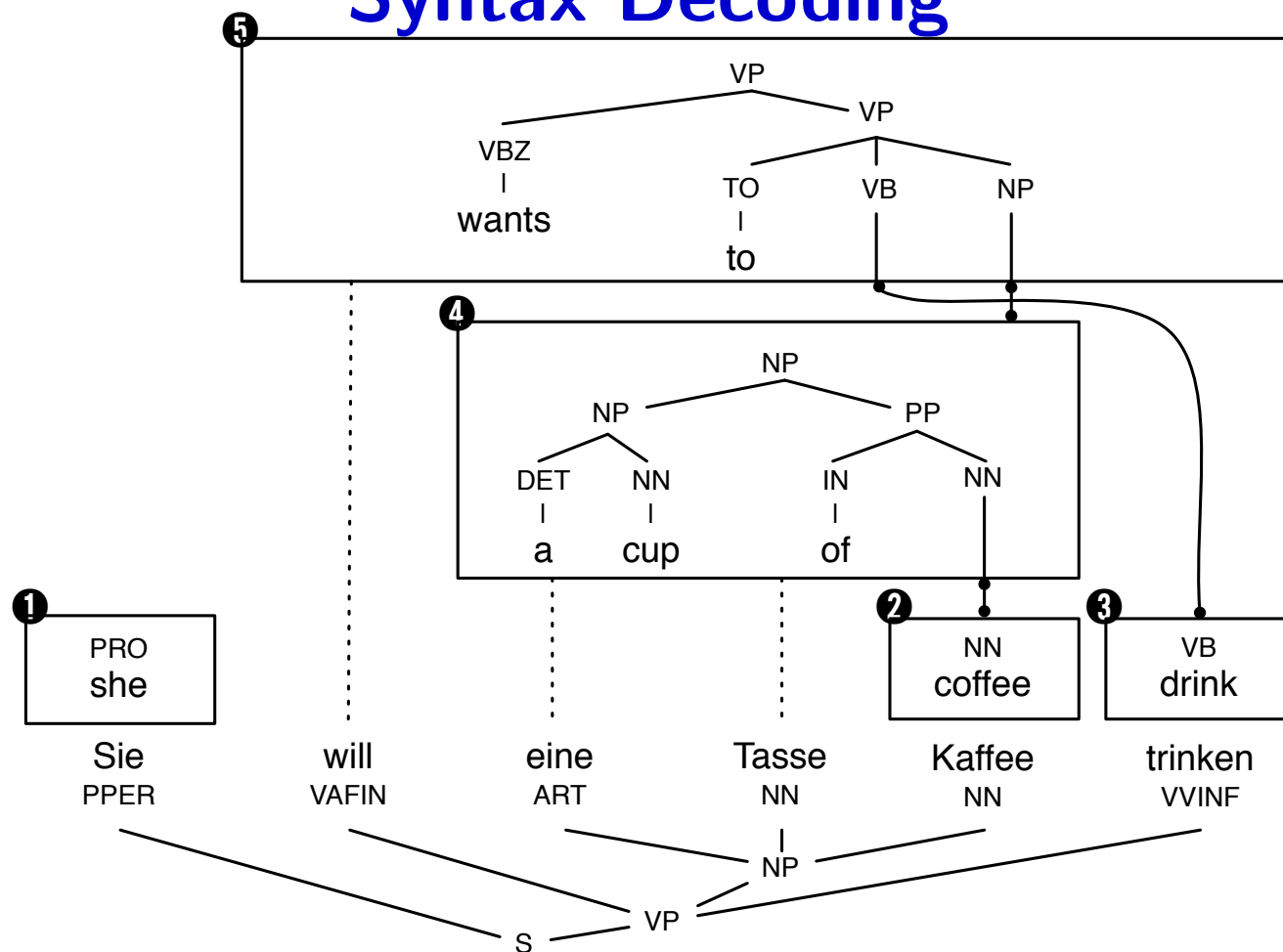
# Syntax Decoding

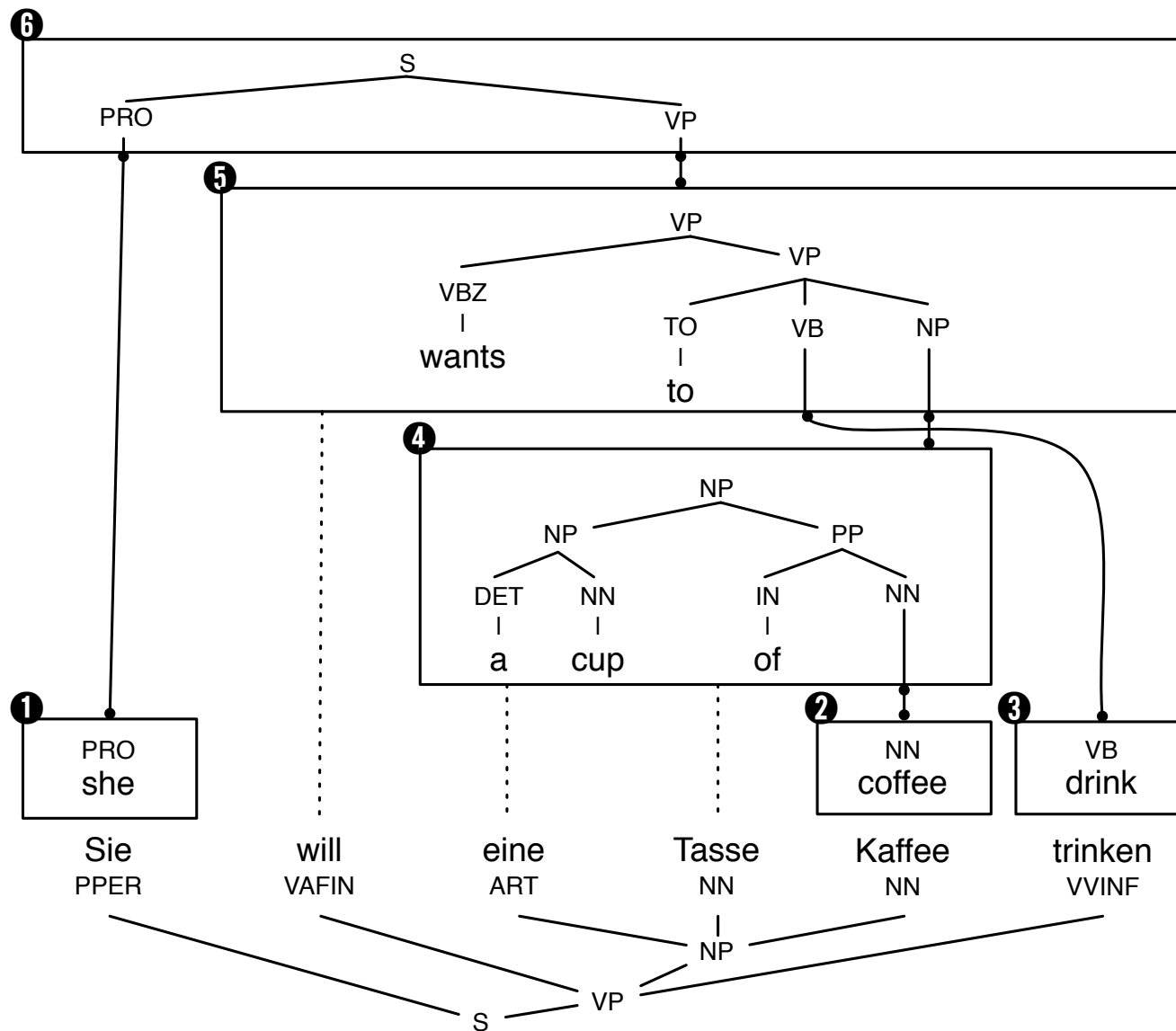


# Syntax Decoding



# Syntax Decoding





# Advanced Topics

- Data and domain adaptation
- Speed vs. quality
- Speed vs. memory use
- Language models
- Instructions to decoder
- Input formats
- Output formats
- Minimum Bayes risk decoding
- Translation models
- Experiment management system



# Hands-On Session

# Advanced Topics

# Advanced Features

- **Data and domain adaptation**
- Speed vs. quality
- Speed vs. memory use
- Language models
- Instructions to decoder
- Input formats
- Output formats
- Minimum Bayes risk decoding
- Translation models
- Experiment management system

# Data

- Parallel corpora → translation model
  - sentence-aligned translated texts
  - translation memories are parallel corpora
  - dictionaries are parallel corpora
- Monolingual corpora → language model
  - text in the target language
  - billions of words easy to handle

# Domain Adaptation

- The more data, the better
- The more in-domain data, the better  
(even in-domain monolingual data very valuable)
- Multiple models
  - train a translation model for each domain corpus
  - train a language model for each domain corpus
  - use all, tune weights for each model
  - alternative: interpolate language model
- Always tune towards target domain

# Advanced Features

- Data and domain adaptation
- **Speed vs. quality**
- Speed vs. memory use
- Language models
- Instructions to decoder
- Input formats
- Output formats
- Translation models
- Experiment management system

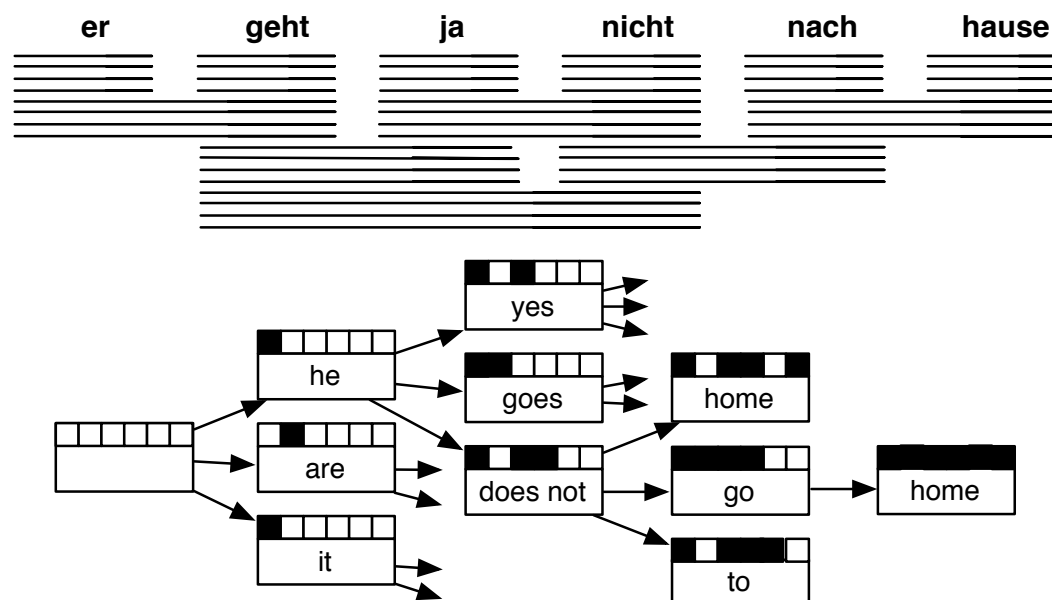
# Speed

- Easy speed-up: multi-threaded decoding

```
--threads NUM
```

- Requires boost library
- Does not currently work for:
  - syntax-based decoding
  - IRSTLM
  - randLM

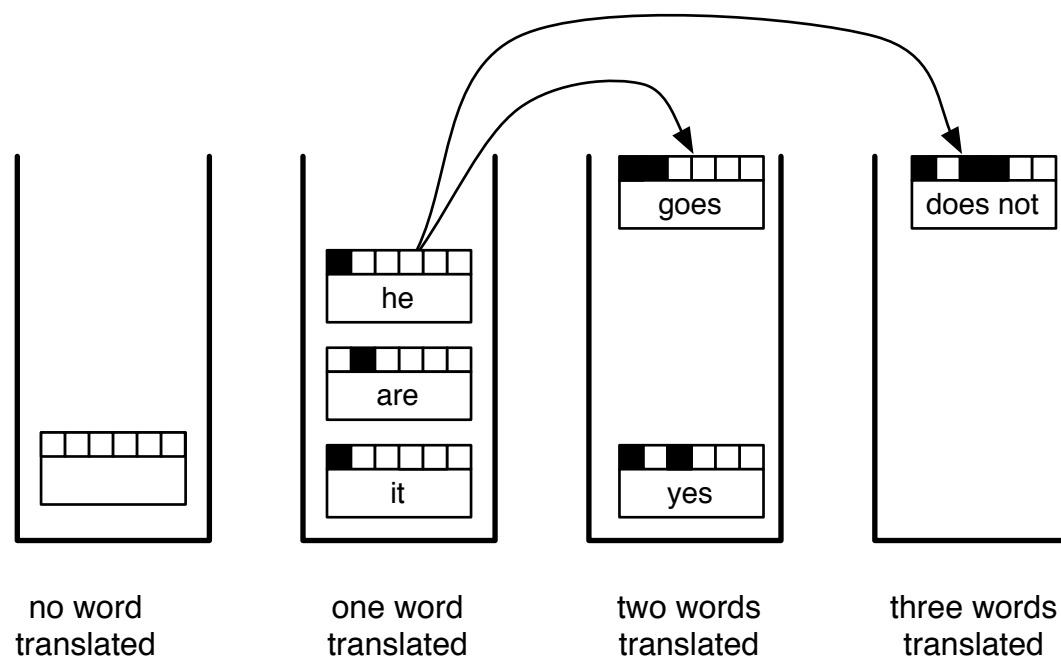
# Speed vs. Quality



- Decoder search creates very large number of partial translations ("hypotheses")
- Decoding time  $\sim$  number of hypotheses created
- Translation quality  $\sim$  number of hypothesis created



# Hypothesis Stacks

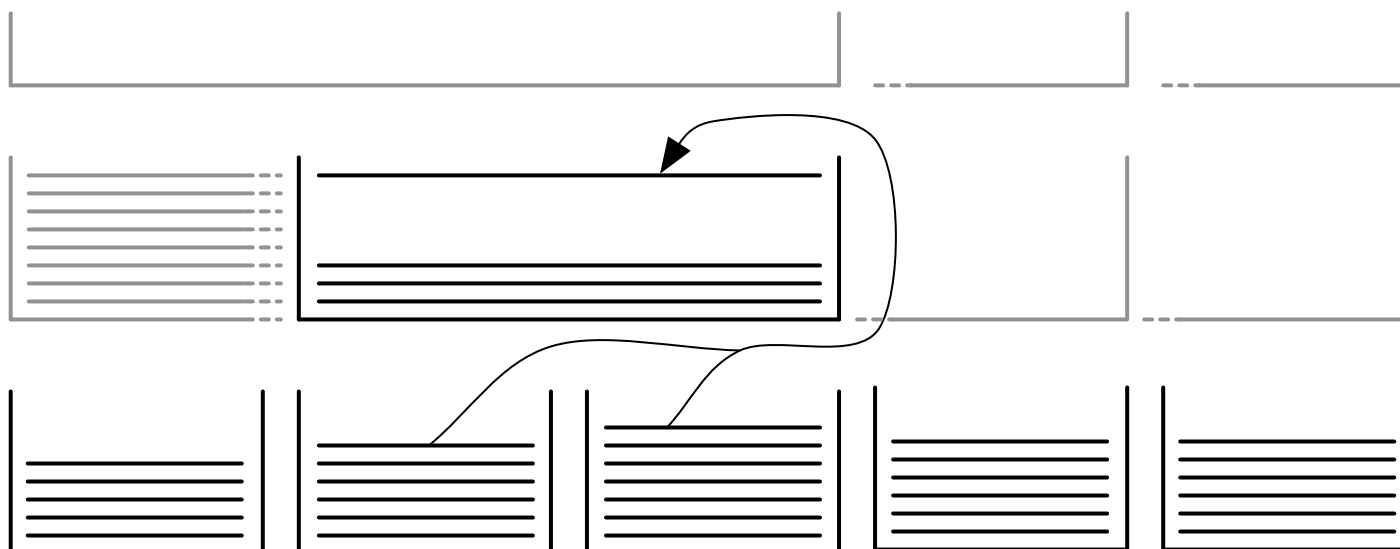


- Phrase-based: One stack per number of input words covered
- Number of hypothesis created =  
sentence length  $\times$  stack size  $\times$  applicable translation options

# Pruning Parameters

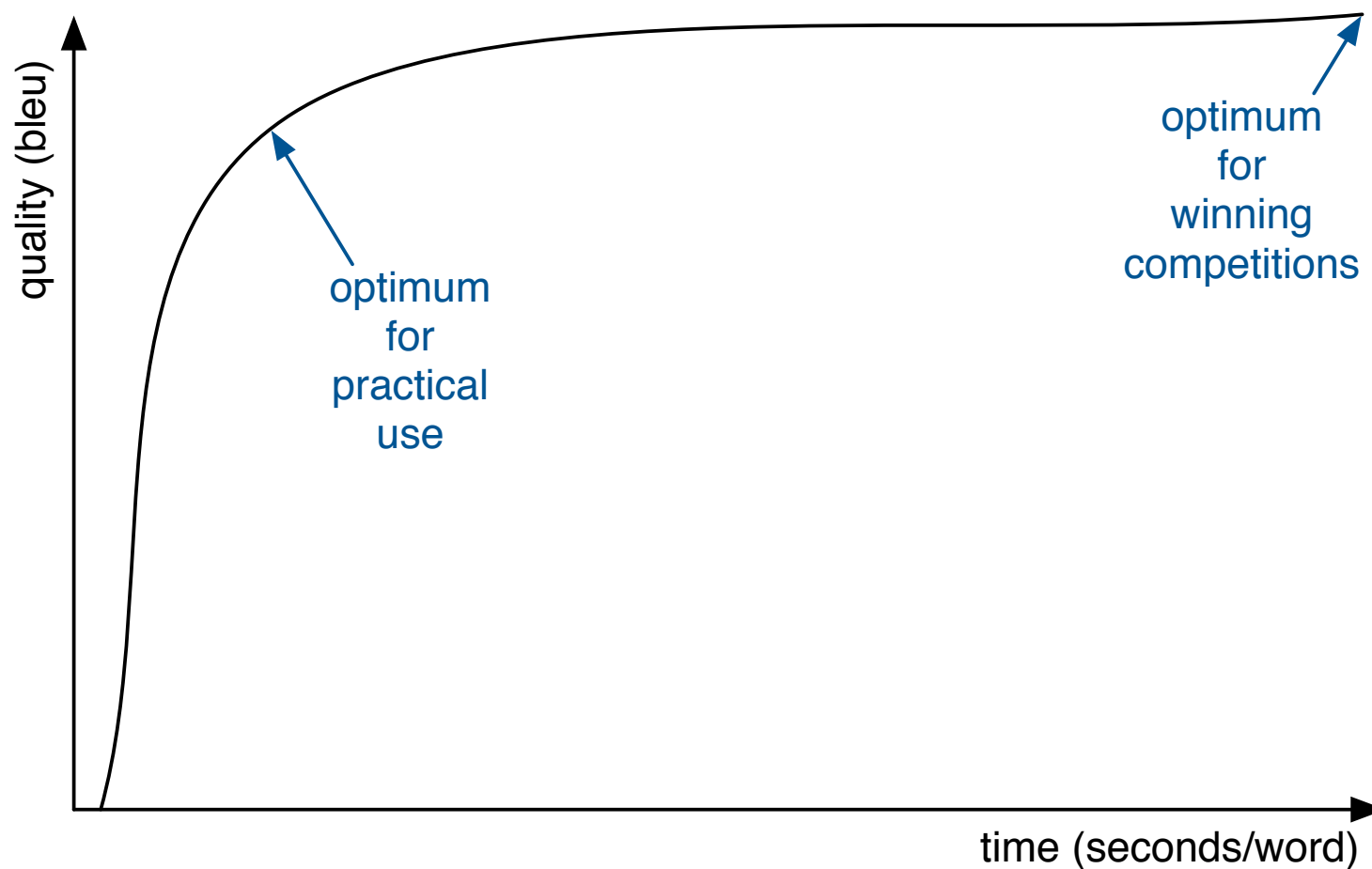
- Regular beam search
  - `--stack NUM` max. number of hypotheses contained in each stack
  - `--ttable-limit NUM` max. num. of translation options per input phrase
  - search time roughly linear with respect to each number
- Cube pruning  
(fixed number of hypotheses are added to each stack)
  - `--search-algorithm 1` turns on cube pruning
  - `--cube-pruning-pop-limit NUM` number of hypotheses added to each stack
  - search time roughly linear with respect to pop limit
  - note: stack size and translation table limit have little impact in speed

# Syntax Hypothesis Stacks



- One stack per input word span
- Number of hypothesis created =  $\text{sentence length}^2 \times \text{number of hypotheses added to each stack}$   
`--cube-pruning-pop-limit NUM` number of hypotheses added to each stack

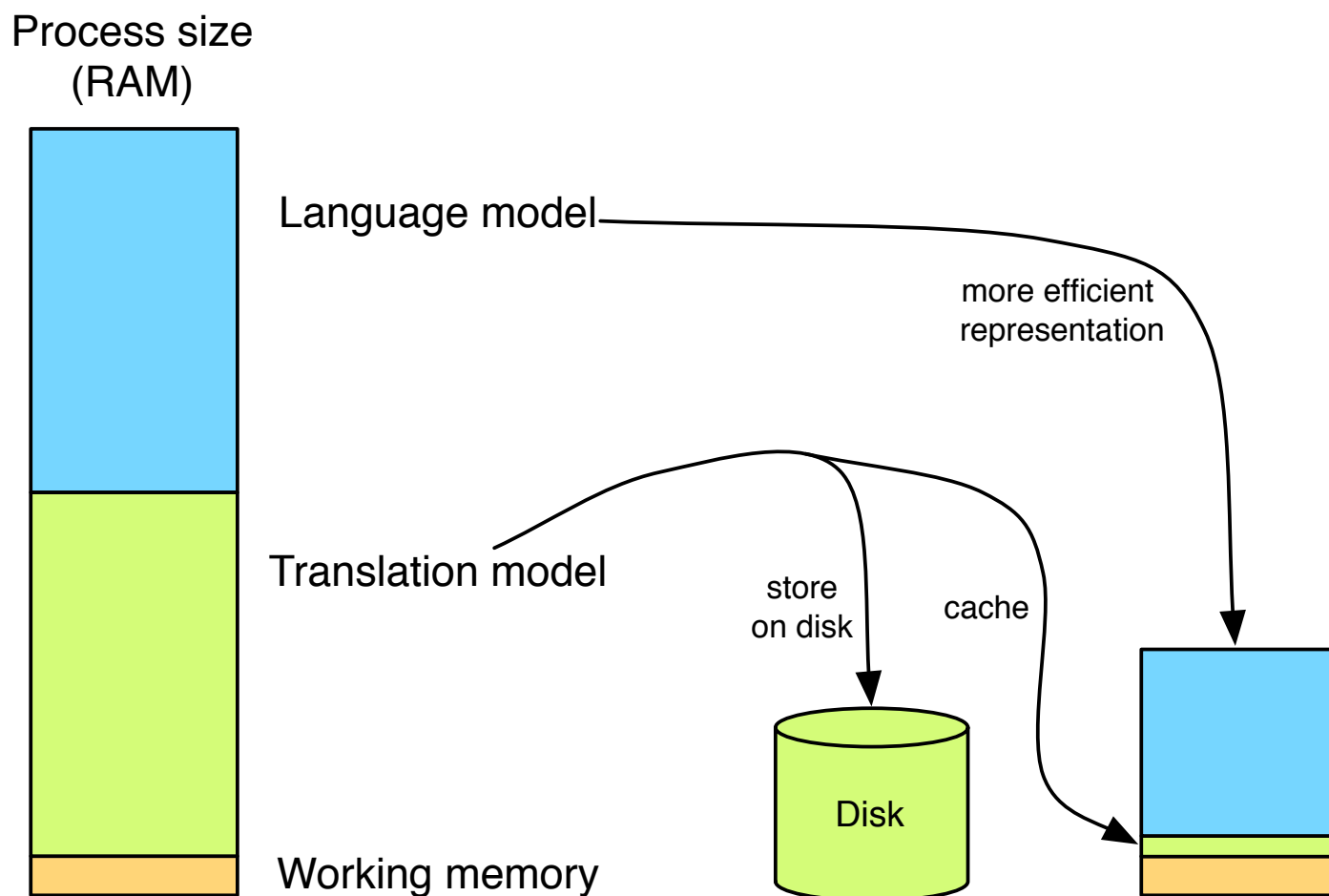
# Trade-Off Speed vs Quality



# Advanced Features

- Data and domain adaptation
- Speed vs. quality
- **Speed vs. memory use**
- Language models
- Instructions to decoder
- Input formats
- Output formats
- Minimum Bayes risk decoding
- Translation models
- Experiment management system

# Speed vs. Memory Use



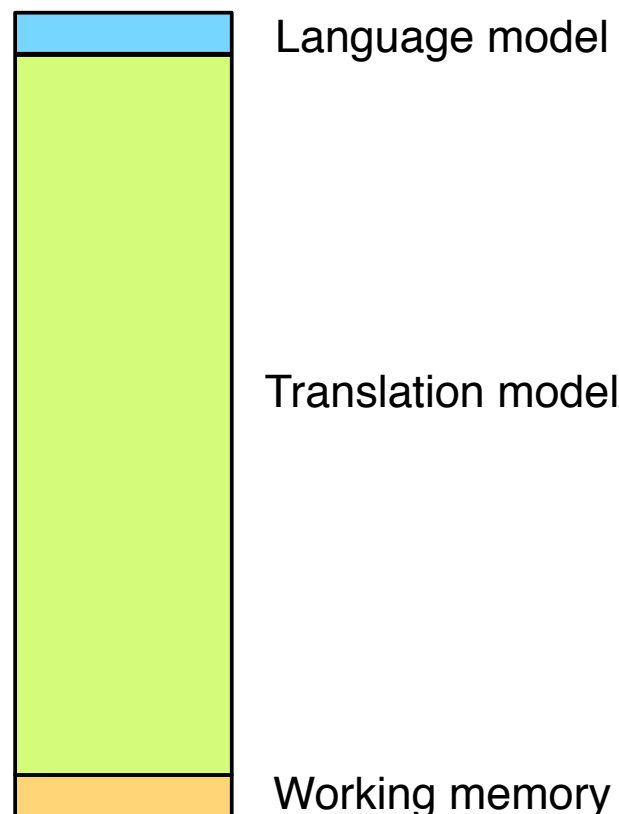
# Speed vs. Memory Use

Typical Europarl file sizes:

- Language model
  - 170 MB (trigram)
  - 412 MB (5-gram)
- Phrase table
  - 11GB
- Lexicalized reordering
  - 9.4GB

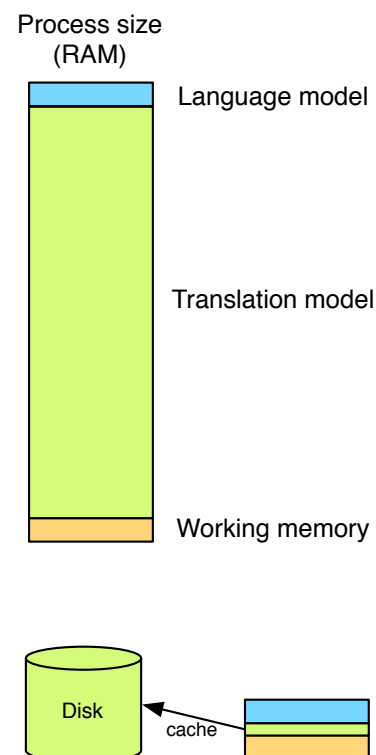
→ total = 20.8 GB

Process size  
(RAM)



# Speed vs. Memory Use

- Load into memory
  - fast decoding
  - large memory usage
  - large load time
- Load-on-demand
  - store indexed model on disk
  - binary format
  - minimal start-up time, memory usage
  - slower decoding





# Speed vs. Memory Use

Phrase Table:

Phrase-based

```
export LC_ALL=C
cat pt.txt | sort | ./processPhraseTable -ttable 0 0 - \
  -nscores 5 -out out.file
```

Hierarchical / Syntax

```
export LC_ALL=C
./CreateOnDiskPt 1 1 5 100 2 pt.txt out.folder
```

Lexical Reordering Table:

```
export LC_ALL=C
processLexicalTable -in r-t.txt -out out.file
```

Language Models (later)

# Speed vs. Memory Use

Change ini file

## Phrase-based

```
[ttable-file]  
1 0 0 5 out.file
```

## Hierarchical / Syntax

```
[ttable-file]  
2 0 0 5 out.folder
```

## Lexical Reordering Table

```
[distortion-file]  
0-0 wbe-msd-bidirectional-fe-allff 6 out.file
```

# Advanced Features



50

- Data and domain adaptation
- Speed vs. quality
- Speed vs. memory use
- **Language models**
- Instructions to decoder
- Input formats
- Output formats
- Minimum Bayes risk decoding
- Translation models
- Experiment management system

# Language Models

- Probability of the output
- Very important in MT, for all SMT models → improve fluency
- Huge amount of training data easy to obtain
  - monolingual
  - can scrape from websites etc.
- But:
  - training takes a long time
  - large memory requirement during decoding
  - large load time
- IRSTLM and RandLM especially designed to tackle large data issues

# IRSTLM



52

- Developed by FBK-irst, Trento, Italy
- Create a binary format which can be read from disk as needed
  - reduces memory but slower decoding
- Quantization of probabilities
  - reduces memory but lose accuracy
  - probability stored in 1 byte instead of 4 bytes

# IRSTLM in Moses

- Compile the decoder with IRSTLM library

```
./configure --with-irstlm=[root dir of the IRSTLM toolkit]
```

- Change ini file to use IRSTLM implementation

```
[lmodel-file]  
1 0 3 file/path
```

# IRSTLM: Training

- Specialized training for large corpora

- parallelization
- reduce memory usage

- Training:

```
build-lm.sh -i "gunzip -c corpus.gz" -n 3  
            -o train.irstlm.gz -k 10
```

- `-n 3` = n-gram order
- `-k 10` = split training procedure into 10 steps

# IRSTLM: Binary Format

- Create binary format:

```
compile-lm language-model.srilm language-model.blm
```

- Load-on-demand:

```
rename file .mm
```



## Randomized language model

- For huge corpora (e.g. 100 billion words)
- Lossy compression
  - Makes false positive mistakes
  - frequency of mistakes can be varied with a parameter
- Typically  $\frac{1}{10}$  size of SRI / IRST language model
- Maybe use as secondary LM to complement conventional LM
  - out-of-domain data scraped from the web
  - high-order n-gram, eg. 6-7 gram

## RandLM: Use in Moses

- Compile the decoder with RandLM library

```
./configure --with-randlm=[root dir of the RandLM toolkit]
```

- Change ini file to use RandLM implementation

```
[lmodel-file]  
0 0 3 /path/to/file    # conventional lm  
5 0 7 /path/to/file    # rand lm
```

# RandLM: Training

- Train from text corpus

```
./buildlm -struct BloomMap -falsepos 8 -values 8 -order 3  
-output-prefix model  
< corpus.txt
```

- Convert SRILM language model

```
./buildlm -struct BloomMap -falsepos 8 -values 8 -order 3  
-output-prefix model  
-input-path lm.srilm -input-type arpa
```

# Advanced Features

- Data and domain adaptation
- Speed vs. quality
- Speed vs. memory use
- Language models
- **Instructions to decoder**
- Input formats
- Output formats
- Minimum Bayes risk decoding
- Translation models
- Experiment management system

# Specifying Translations with XML

- Translation tables for numbers?

$f$	$e$	$p(f e)$
2003	2003	0.7432
2003	2000	0.0421
2003	year	0.0212
2003	the	0.0175
2003	...	...

- Instruct the decoder with XML instruction

the revenue for <num translation="2003"> 2003 </num> is higher than ...

- Deal with different number formats

er erzielte <num translation="17.55"> 17,55 </num> Punkte .

# Walls and Zones

- Specification of reordering constraints
- Zone  
sequence to be translated without reordering with outside material
- Wall  
hard reordering constraint, no words may be reordered across
- Local wall  
wall within a zone, not valid outside zone

# Walls and Zones: Examples

- Requiring the translation of quoted material as a block

He said <zone> " yes " </zone> .

- Hard reordering constraint

Number 1 : <wall/> the beginning .

- Local hard reordering constraint within zone

A new plan <zone> ( <wall/> maybe not new <wall/> ) </zone> emerged .

- Nesting

The <zone> " new <zone> ( old ) </zone> " </zone> proposal .

# Preserving Markup

- How do you translate this:

`<h1>My Home Page</h1>`  
`I really like to <b>eat</b> chicken!`

- Solution 1: XML translations, walls and zones

```
<x translation="<h1>" /> <wall/> My Home Page <wall/>  
<x translation="</h1>" />
```

```
I really like to <zone><x translation="<b>" /> <wall/> eat <wall/>  
<x translation="</b>" /> </zone> chicken !
```

(note: special XML characters like `<` and `>` need to be escaped)



## Preserving Markup

- Solution 2: Handle markup externally

- track word positions and their markup

I	really	like	to	<b>eat</b>	chicken	!
1	2	3	4	5	6	7
-	-	-	-	<b>	-	-

- translate without markup

I really like to eat chicken !

- keep word alignment to source

Ich	esse	wirklich	gerne	Hühnchen	!
1	5	2	3-4	6	7

- re-insert markup

Ich <b>esse</b> wirklich gerne Hühnchen!

# Advanced Features

- Data and domain adaptation
- Speed vs. quality
- Speed vs. memory use
- Language models
- Instructions to decoder
- **Input formats**
- Output formats
- Minimum Bayes risk decoding
- Translation models
- Experiment management system

## Example: Misspelt Words

- Misspelt sentence:

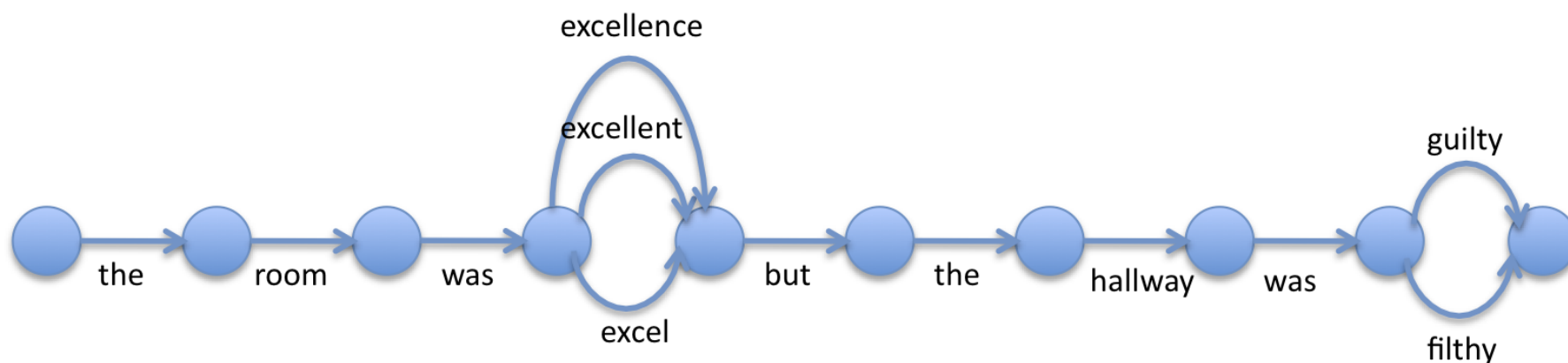
The room was \*exellent but the hallway was \*filty .

- Strategies for dealing with spelling errors:
  - Create correct sentence with correction
    - ✗ problem: if not corrected properly, adds more errors
  - Create many sentences with different corrections
    - ✗ problem: have to decode each sentence, slow

# Confusion Network

The room was \*excellent but the hallway was \*filthy .

Input to decoder:



Let the decoder decide

## Example: Diacritics

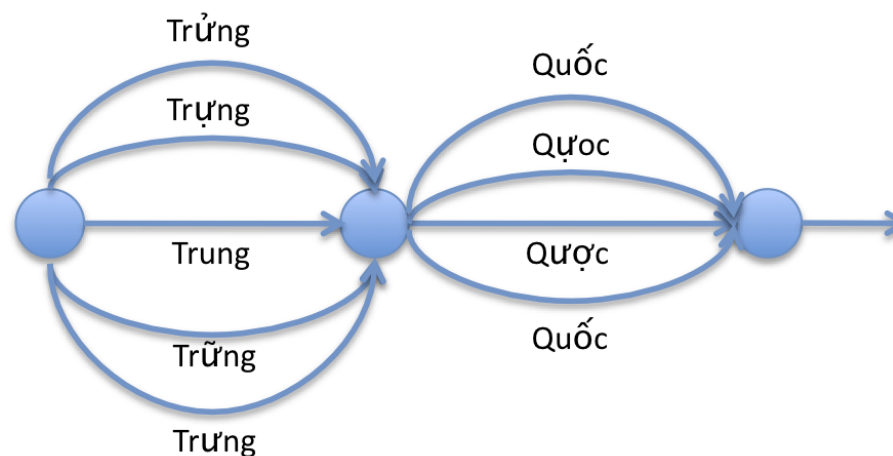
- Correct sentence

Trung Quốc cảnh báo Mỹ về luật tiền tệ

- Something a non-native person might type

Trung Quoc canh bao My ve luat tien te

- Confusion network



# Confusion Network Specification

Argument on command line

```
./moses -inputtype 1
```

Input to moses

```
the 1.0  
room 1.0  
was 1.0  
excel 0.33 excellent 0.33 excellence 0.33  
but 1.0  
the 1.0  
hallway 1.0  
was 1.0  
guilty 0.5 filthy 0.5
```

# Lattice

## Example: Chinese Word Segmentation

- Unsegmented sentence

硬质合金号称"工业牙齿"

- Incorrect segmentation

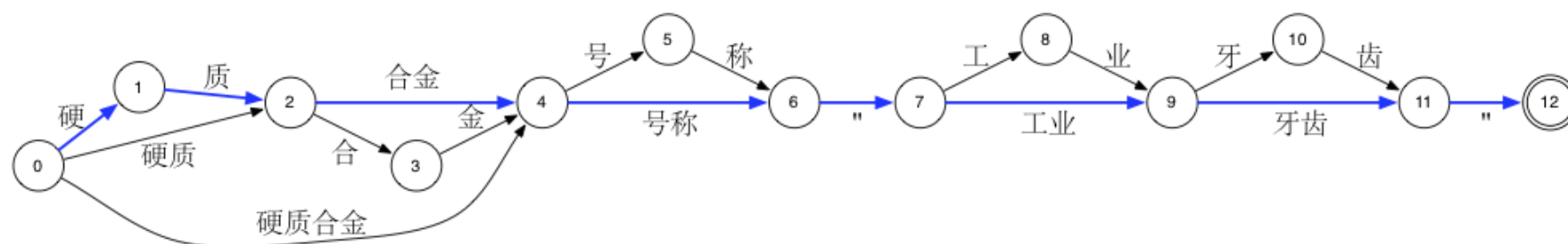
硬质 合 金 号称 " 工 业牙 齿 "

- Correct segmentation

硬 质 合金 号称 " 工业 牙齿 "

# Lattice

Input to decoder:



Let the decoder decide

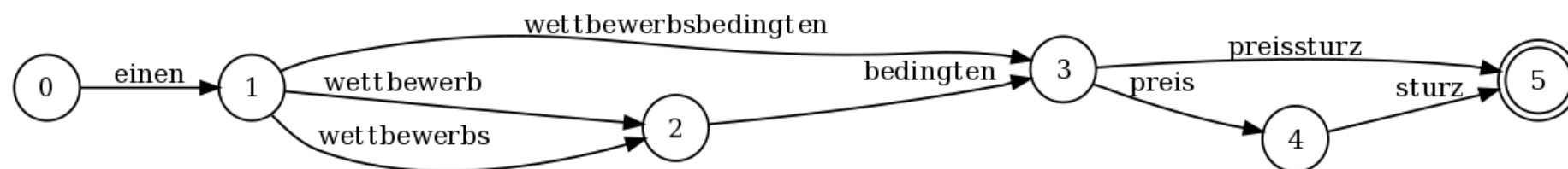


## Example: Compound Splitting

- Input sentence

einen wettbewerbsbedingten preissturz

- Different compound splits



- Let the decoder decide

# Lattice Specification

## Command line argument

```
./moses -inputtype 1
```

## Input to Moses (PLF format - Python Lattice Format)

```
(  
  (  
    ('einen', 1.0, 1),  
  ),  
  (  
    ('wettbewerbsbedingen', 0.5, 2),  
    ('wettbewerbs', 0.25, 1),  
    ('wettbewerb', 0.25, 1),  
  ),  
  (  
    ('bedingen', 1.0, 1),  
  ),  
  (  
    ('preissturz', 0.5, 2),  
    ('preis', 0.5, 1),  
  ),  
  (  
    ('sturz', 1.0, 1),  
  ),  
)
```

# Advanced Features

- Data and domain adaptation
- Speed vs. quality
- Speed vs. memory use
- Language models
- Instructions to decoder
- Input formats
- **Output formats**
- Minimum Bayes risk decoding
- Translation models
- Experiment management system

# N-Best List

- Input

es gibt verschiedene andere meinungen .

- Best Translation

there are various different opinions .

- Next nine best translations

there are various other opinions .  
there are different different opinions .  
there are other different opinions .  
we are various different opinions .  
there are various other opinions of .  
it is various different opinions .  
there are different other opinions .  
it is various other opinions .  
it is a different opinions .

# Uses of N-Best Lists

- Let the translator choose from possible translations
- Reranker
  - add more knowledge sources
  - can take global view
  - coherency of whole sentence
  - coherency of document
- Used to tune component weights

# N-Best Lists in Moses

Argument to command line

```
./moses -n-bestlist n-best.file.txt [distinct] 100
```

Output

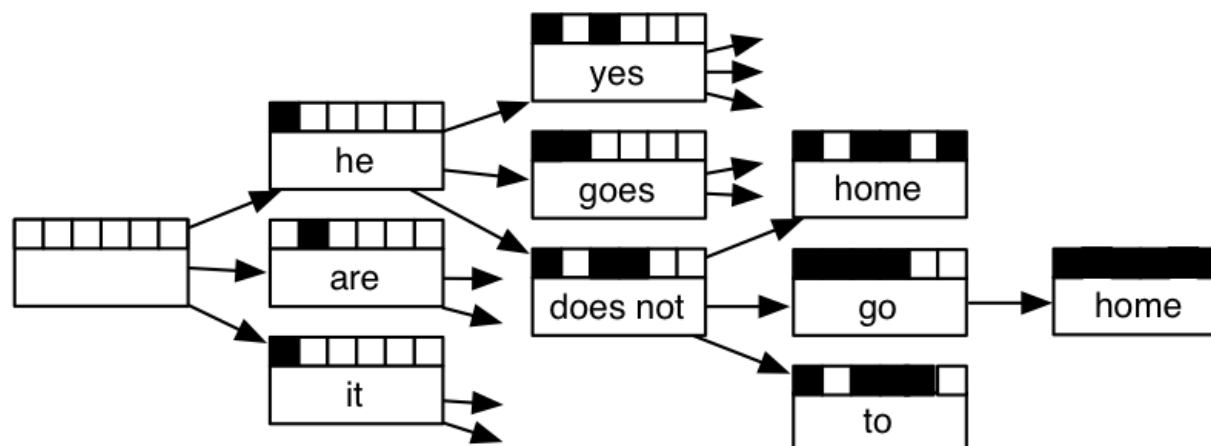
```
0 ||| there are various different opinions . ||| d: 0 lm: -21.6664 w: -6 ... ||| -113.734
0 ||| there are various other opinions . ||| d: 0 lm: -25.3276 w: -6 ... ||| -114.004
0 ||| there are different different opinions . ||| d: 0 lm: -27.8429 w: -6 ... ||| -117.738
0 ||| there are other different opinions . ||| d: -4 lm: -25.1666 w: -6 ... ||| -118.007
0 ||| we are various different opinions . ||| d: 0 lm: -28.1533 w: -6 ... ||| -118.142
0 ||| there are various other opinions of . ||| d: 0 lm: -33.7616 w: -7 ... ||| -118.153
0 ||| it is various different opinions . ||| d: 0 lm: -29.8191 w: -6 ... ||| -118.222
0 ||| there are different other opinions . ||| d: 0 lm: -30.426 w: -6 ... ||| -118.236
0 ||| it is various other opinions . ||| d: 0 lm: -32.6824 w: -6 ... ||| -118.395
0 ||| it is a different opinions . ||| d: 0 lm: -20.1611 w: -6 ... ||| -118.434
```

# Search Graph

- Input

er geht ja nicht nach hause

- Return internal structure from the decoder



- Encode millions of other possible translations  
(every path through the graph = 1 translation)

# Uses of Search Graphs

- Let the translator choose
  - Individual words or phrases
  - 'Suggest' next phrase
- Reranker
- Used to tune component weights
  - More difficult than with n-best list

[1] New probe into US attorney affair >>  
 Neuer Vorstoß in den USA Anwalt neue Affäre sonde (9 edits)

neue sonde

enter in

new	probe	into	US	attorney	affair
neue	Sonde	in		Anwalt	die
die	testet	In	die	Staatsanwalt	Affäre
die	prüfen	In	In	Anwälte	die
der	Vorstoß	In	die	Testamentsvollstreckers	sie
eine	auszuforschen	In	die	Vollmachten	Angelegenheit
neuer	prüfen	auch	In	Anwalt	um
die	prüfen	In	der		Sache
das	prüfen	zu	amerikanische		haben
neu	prüfen	In	der		Geschichte
In		nach	die		das



# Search Graphs in Moses

Argument to command line

```
./moses -output-search-graph search-graph.file.txt
```

Argument to command line

```
0 hyp=0 stack=0 forward=36 fscore=-113.734
0 hyp=75 stack=1 back=0 score=-104.943 ... covered=5-5 out=.
0 hyp=72 stack=1 back=0 score=-8.846 ... covered=4-4 out=opinions
0 hyp=73 stack=1 back=0 score=-10.661 ... covered=4-4 out=opinions of
```

- hyp - hypothesis id
- stack - how many words have been translated
- score - total weighted score
- covered - which words were translated by this hypothesis
- out - target phrase

# Advanced Features

- Data and domain adaptation
- Speed vs. quality
- Speed vs. memory use
- Language models
- Instructions to decoder
- Input formats
- Output formats
- **Minimum Bayes risk decoding**
- Translation models
- Experiment management system

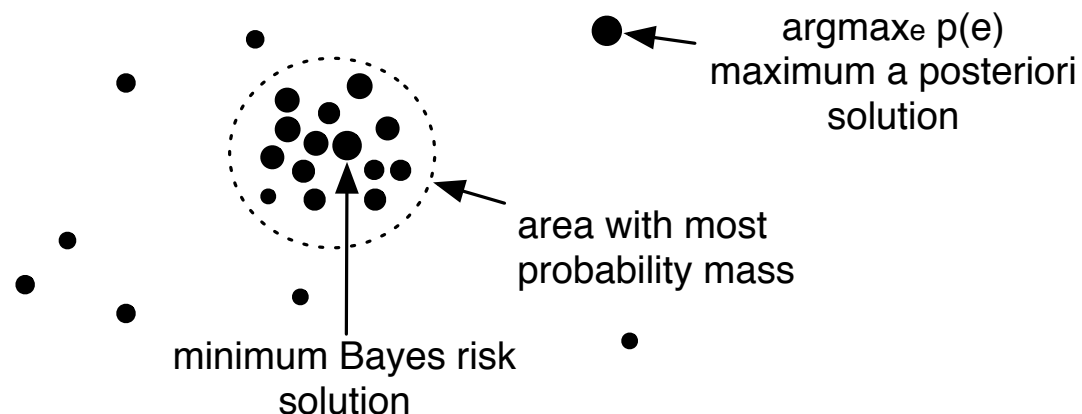
# Minimum Bayes Risk Decoding

- Normal (MAP) decoding:

$$\hat{t} = \operatorname{argmax}_t p(t|s)$$

- MBR decoding:

$$\hat{t} = \operatorname{argmax}_t \sum_{t' \in T} p(t'|s) \times \text{bleu}(t', t)$$



# Minimum Bayes Risk Decoding

- Set of translations  $t' \in T$

$$\hat{t} = \operatorname{argmax}_t \sum_{t' \in T} p(t'|s) \times \operatorname{bleu}(t', t)$$

- Using n-best list:

```
moses -f moses.ini -i in.txt -mbr
```

- Using lattice:

```
lmbrgrid ... -f moses.ini -i input.txt
```

# Advanced Features

- Data and domain adaptation
- Speed vs. quality
- Speed vs. memory use
- Language models
- Instructions to decoder
- Input formats
- Output formats
- Minimum Bayes risk decoding
- **Translation models**
- Experiment management system

# Phrase-Based Model

- Advantages

- fast: under half a second per sentence for fast configuration
- low-memory requirement
  - \* 200-300MB for lowest configuration
  - \* suitable for netbooks and mobile devices
- outperform more complicated models for many language pairs
  - \* especially for related languages pairs

- Command line

```
./moses -f moses.ini -i in.txt > out.txt
```

- Output

```
there are various different opinions .
```

# Hierarchical Models

## Advantages

- able to model non-contiguous phrases
  - ne..pas → not
- low-memory requirement
  - 200-300MB for lowest configuration
  - suitable for netbooks and mobile devices
- outperform phrase-based models when translating between widely different languages
  - Chinese-English consistently better with hierarchical model
  - better at medium range re-ordering
- Linguistically motivated

## Disadvantages

- slower
  - 0.5 - 2 sec for fastest configuration
- more memory requirement
  - 1-2GB ram
- more disk usage
  - translation model  $\times 10$  larger than phrase-based

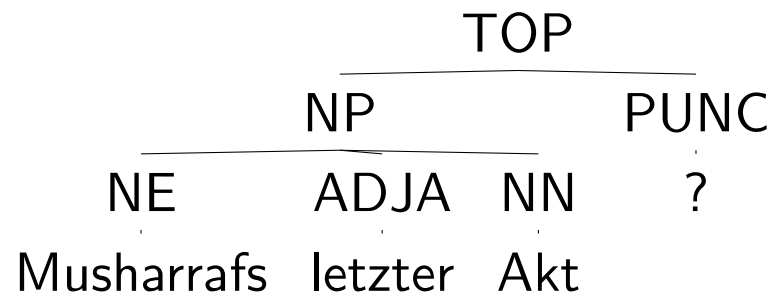
**Command line** `./moses-chart -f moses.ini -i in.txt > out.txt`

# Syntax Models

- Hierarchical model + use of syntactic information (constituency parser, chunkers)
- Advantage
  - Can use outside linguistic information
  - promises to solve important problems in SMT, eg. long-range reordering
- Disadvantages
  - difficult to get right
  - for many language pairs still worse than phrase-based and hierarchical models
  - need syntactic parse information
    - \* unreliable
    - \* available only for some languages
    - \* not designed for machine translation



# Moses Tree Representation



```

- <tree label="TOP">
  - <tree label="NP">
    <tree label="NE"> Musharrafs </tree>
    <tree label="ADJA"> letzter </tree>
    <tree label="NN"> Akt </tree>
  </tree>
  <tree label="PUNC"> ? </tree>
</tree>
  
```

# Phrase-Based Model Training

- Command line

```
train-model.perl ...
```

- Model

```
Bndnisse ||| alliances ||| 1 1 1 1 2.718 ||| ||| 1 1  
General Musharraf betrat am ||| general Musharraf appeared on ||| 1 1 1 1 2.718 ||| ||| 1 1
```

# Hierarchical Model Training

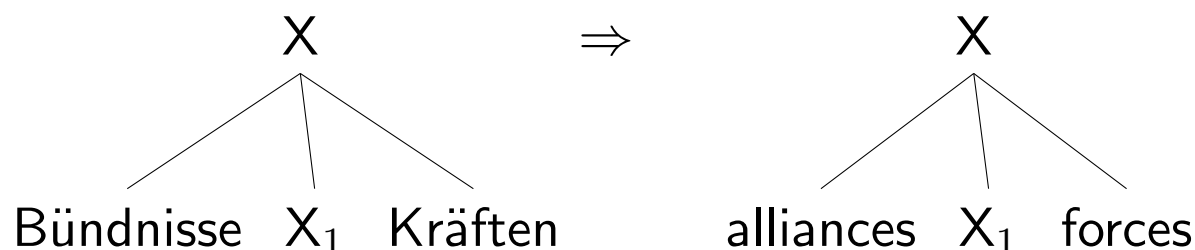
- Command line

```
train-model.perl ... -hierarchical
```

- Example rule from model

```
Bündnisse [X][X] Kräften [X] ||| alliances [X][X] forces [X] ||| 1 1 1 1 2.718 ||| 1-1 ||| 0.0526316 0.0526316
```

- Visualization of rule



# Tree-to-String Model Training

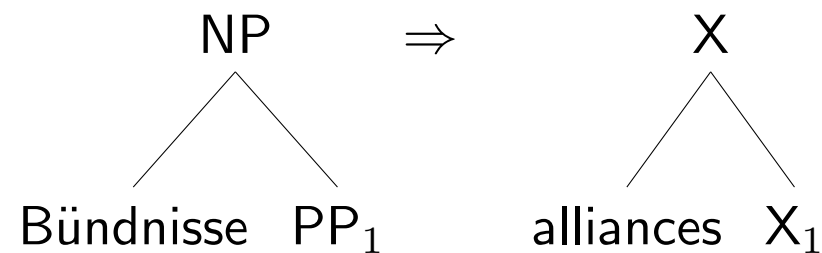
- Command line

```
train-model.perl ... -source-syntax
```

- Example rule from model

```
Bündnisse [PP][X] [NP] ||| alliances [PP][X] [X] ||| 1 1 1 1 2.718 ||| 1-1 ||| 1 1
```

- Visualization of rule



# String to Tree Model Training

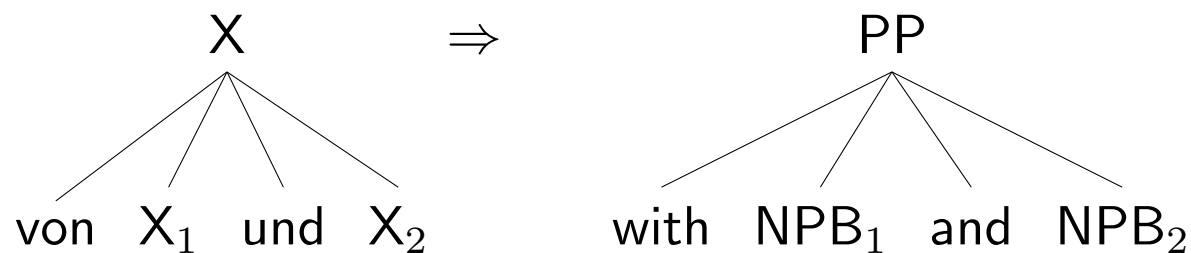
- Command line

```
train-model.perl ... -target-syntax
```

- Example rule from model

```
von [X][NPB] und [X][NPB] [X] ||| with [X][NPB] and [X][NPB] [PP] ||| ...
```

- Visualization of rule



# Tree-to-Tree Model Training

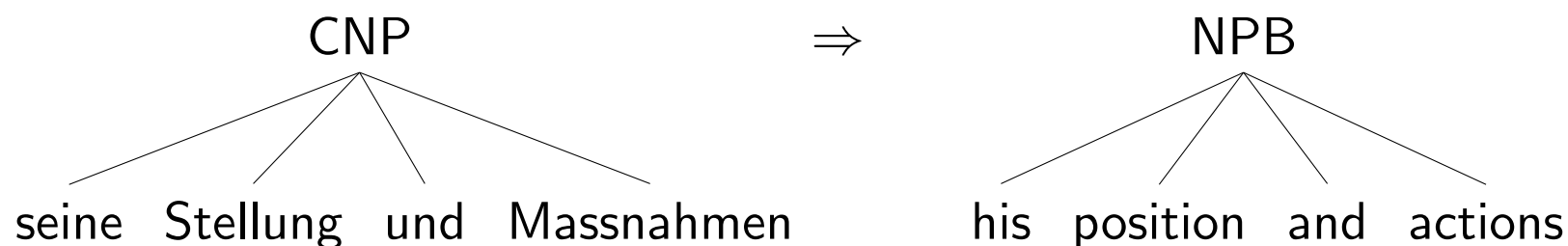
- Command line

```
train-model.perl ... -source-syntax -target-syntax
```

- Example rule from model

```
seine Stellung und Manahmen [CNP] ||| his position and actions [NPB] ||| ...
```

- Visualization of rule



# Syntax Models Decoding in Moses

- String-to-string (hierarchical) or string-to-tree

```
./moses-chart -f moses.ini -i in.txt > out.txt
```

- Tree-to-string or tree-to-tree

```
./moses-chart -f moses.ini -i in.txt -inputtype 3 > out.txt
```

# Advanced Features

- Data and domain adaptation
- Speed vs. quality
- Speed vs. memory use
- Language models
- Instructions to decoder
- Input formats
- Output formats
- Minimum Bayes risk decoding
- Translation models
- **Experiment management system**



# Running Experiments

Execute a lot of scripts

```
tokenize < corpus.en > corpus.en.tok  
lowercase < corpus.en.tok > corpus.en.lc  
...  
mert.perl ....  
moses ...  
mteval-v13.pl ...
```

Change a part of the process, execute everything again

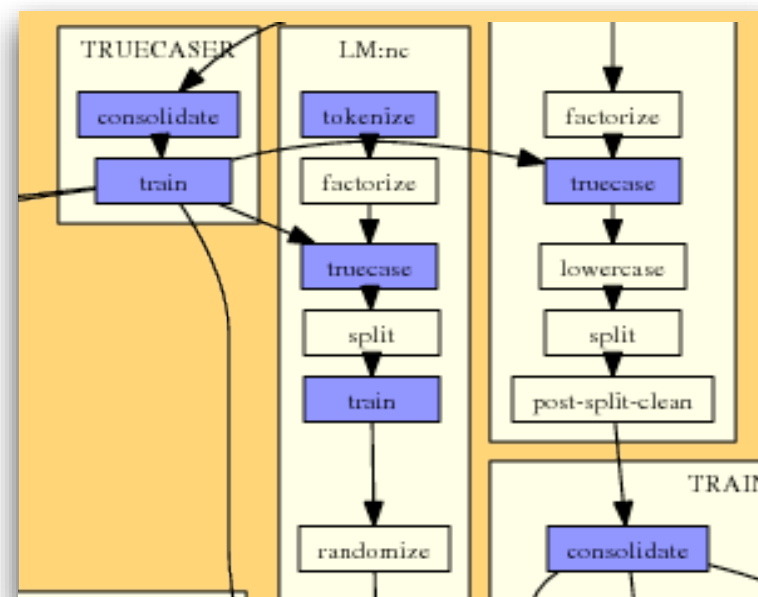
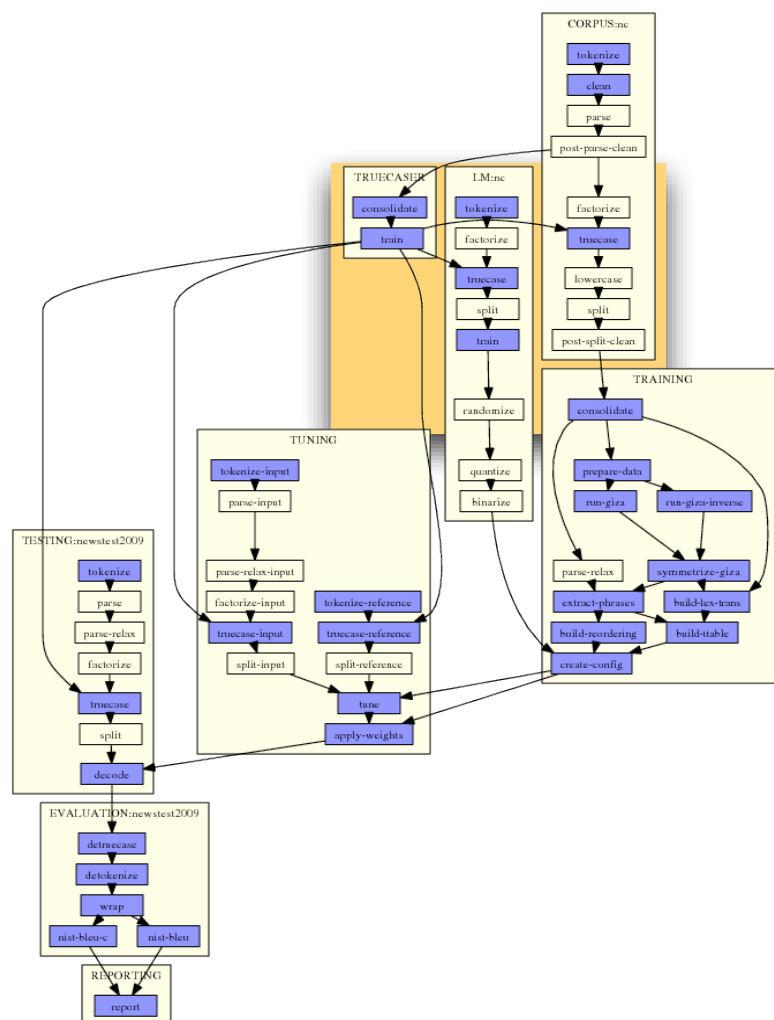
```
tokenize < corpus.en > corpus.en.tok  
lowercase < corpus.en.tok > corpus.en.lc  
...  
mert.perl ....  
moses ...  
mteval-v13.pl ...
```

# Experiment Management System



- One configuration file for all settings: record of all experimental details
- Scheduler of individual steps in pipeline
  - automatically keeps track of dependencies
  - on single machine, multi-core machines, GridEngine clusters
  - parallel execution
  - crash detection
  - automatic re-use of prior results
- Fast to use
  - set up a new experiments in minutes
  - set up a variation of an experiment in seconds

Workflow automatically  
generated by  
experiment.perl



# How does it work?

- Write a configuration file (typically by adapting an existing file)

- Execute:

```
experiment.perl -config config
```

# Web Interface

100



## All Experimental Setups

ID	User	Task	Directory
<a href="#">97</a>	pkoehn	Acquis Truecased	/group/project/statmt2/pkoehn/acquis-truecase
<a href="#">96</a>	pkoehn	Chinese-English AGILE 2008	/group/project/statmt2/pkoehn/agile08-chinese
<a href="#">95</a>	miles	Randlm testing	/group/project/statmt7/miles/experiments /ep-enfr/work
<a href="#">94</a>	joseph	Proj2008 Impl.Adapted experiment(fr-en)for News Comm.	/group/project/statmt2/joseph/experimentJo/task6
<a href="#">93</a>	joseph	Proj2008 Impl.Baseline experiment(fr-en)for News Comm.	/group/project/statmt2/joseph/experimentJo/task5
<a href="#">92</a>	jschroe1	FR-EN System Combination Components	/group/project/statmt9/josh/experiments /fr-syscomb/work

List of experiments

# List of Runs

## Task: WMT10 German-English (pkoehn)

[Wiki Notes](#) | [Overview of experiments](#) | [/fs/bragi2/pkoehn-experiment/wmt10-de-en](#)

compare	ID	start	end	avg	newstest2009		newstest2010	
<input type="checkbox"/> <a href="#">cfglparlimg</a>	<a href="#">[1042-16]</a> 11+analysis	16 May	16 May	BLEU-c: 21.74 BLEU: 22.91	<a href="#">21.03</a> <a href="#">(1.002)</a> <a href="#">22.30</a> <a href="#">(1.002)</a>	<a href="#">A</a> <input type="checkbox"/>	<a href="#">22.45</a> <a href="#">(1.041)</a> <a href="#">23.51</a> <a href="#">(1.041)</a>	<a href="#">A</a> <input type="checkbox"/>
<input type="checkbox"/> <a href="#">cfglparlimg</a>	<a href="#">[1042-15]</a> 11+Internal emplus test set	21 Apr	crashed	-	-		-	
<input type="checkbox"/> <a href="#">cfglparlimg</a>	<a href="#">[1042-14]</a> 9+interpolated-tm.lm-weighted	21 Feb	21 Feb 9: 0.239258 -> 0.239296	-	<a href="#">20.81</a> <a href="#">(1.003)</a> <a href="#">22.06</a> <a href="#">(1.003)</a>	<a href="#">A</a> <input type="checkbox"/>	-	
<input type="checkbox"/> <a href="#">cfglparlimg</a>	<a href="#">[1042-13]</a> 9+only-ep	21 Feb	21 Feb 13: 0.235046 -> 0.235053	-	<a href="#">20.42</a> <a href="#">(1.002)</a> <a href="#">21.69</a> <a href="#">(1.002)</a>	<a href="#">A</a> <input type="checkbox"/>	-	
<input type="checkbox"/> <a href="#">cfglparlimg</a>	<a href="#">[1042-12]</a> 9+only-nc	21 Feb	21 Feb 7: 0.222237 ->	-	<a href="#">18.96</a> <a href="#">(1.002)</a> <a href="#">20.16</a>	<a href="#">A</a> <input type="checkbox"/>	-	

# Analysis: Basic Statistics

Coverage			Phrase Segmentation				
model	corpus			1	2	3	4+
0	2047 (3.1%)	1708 (2.6%)	1 to	26897 (40.7%)	2145 (3.2%)	278 (0.4%)	90 (0.1%)
1	738 (1.1%)	518 (0.8%)	2 to	4144 (6.3%)	14414 (21.8%)	2518 (3.8%)	432 (0.7%)
2-5	1483 (2.2%)	818 (1.2%)	3 to	639 (1.0%)	3522 (5.3%)	4821 (7.3%)	1272 (1.9%)
6+	61745 (93.5%)	62969 (95.4%)	4+ to	158 (0.2%)	855 (1.3%)	1693 (2.6%)	2135 (3.2%)
by token / <a href="#">by type</a> / <a href="#">details</a>			by word / <a href="#">by phrase</a>				

- Basic statistics
  - n-gram precision
  - evaluation metrics
  - coverage of the input in corpus and translation model
  - phrase segmentations used



# Analysis: Unknown Words

grouped by frequency in test set

## unknown words

18 Eatonville	<b>4:</b>	<b>3:</b> Anmil,	<b>2:</b> Abfertigungen,	<b>1:</b> -Ach, -Minister, -Pakets, -weiss, .docx, .pptx, .xlsx, 1,45,
16 Hurston	Eatonvilles,	Atlasz, BR23C,	Albums, Alondra,	1.106,55, 1.983,73, 10.365,45, 10.579, 10.809,25, 106,85,
12 Barrick	Együtt,	BSA, Bayón,	Andoh, Anm., Armiñon,	11,9, 11.743,61, 12.595,75, 14,2, 14,7, 145,29, 16,8, 17,9,
12 Hema	Garver,	Biztos, Bt.,	Ashford, BZÖ, Baloldal,	18,6, 18.286,90, 1802, 1834, 1880ern, 1920ern, 1925,
12 Stewards	Harmadik,	Butch, Casado,	Bani, Baugesellschaften,	19252008, 199,61, 2,178, 2,37, 2.400, 26,3, 270.000, 29,2,
11 Gebrselassie	Hurstons,	Dal, Embraer,	Bedienkomfort, Bento,	3,30, 3,632, 3,827, 3.0.0, 4,161, 4,357, 42,2, 43,4, 499,
10 Flamenco	Jobb, Jol,	FT, Faymann,	Bentos, Bingleys, Bojen,	49sten, 5.839, 506,43, 6,98, 684,81, 729,700, 75,5, 777,68,
10 Mango	Jos, Jövőért,	Fiatal, Gregg,	Bowens, Bowery, Boyd,	8,25, 8,81, 9,14, 99,80, AAC, ADQ, ART, Aareal,
9 Glitter	Kovalev,	Gélineau, HSV,	Bringley, Browser,	Abbremsens, Abhöraktion, Absenzen, Abwesenheiten,
9 ÚOHS	Kreuer,	Hanzelka,	Bělohávek, CBGB,	Abwiegen, Abwärtssog, Achronot, Actor, AdSense,
9 ČTÚ	Lados,	Illhäusern, Iván,	Carci, Cera, Charts,	AdWords, Aday, Adobe, Adressverzeichnisses, Adwards,
8 Coles	Mercandelli,	Jansen, Jančura,	Chemical, Chigi,	Adélar, Agazio, Akku, Akron, Aktuálně.cz, Alameda,
8 Deka	Stehplätze,	Joanne,	Cineast, Comics,	Alatriste, Alcolock, Aleš, Alhambra, Alleinregierer,
8 Garci	Tauro,	Kemrová, Kid,	Commerzbank, Coppola,	Amazonengebiet, Amil, Aminei, Amministrazione, Amway,
8 ITV	Tórtola,	Llamazares,	Corker, Cowon, DF,	Andalusierin, Andik, Android, Anděl, Angeklagtem, Ansa,
	Zenobia,	Loafs, Mangas,	Dinkins, Download,	Anthologie, Antiasthmatica, Apnoe, Aquel, Arabija,
	fon,	Medikamentes,	Drehbewegung,	Arbeitenehmers, Arcandor, Arriaga, Asiana, Askale,
	Évezredért,	Mobil.cz,	Drzewiecki, Drápal,	Astronomen, Aufeislegen, Augäpfel, Ausdrückstärke,
	Ózd	Mutual,	Düsseldorfer, Ella,	Ausführungs-, Ausgeruhter, Ausscheidungsspiele,



# Analysis: Output Annotation

104



[0.2152] This time was the reason for the collapse on Wall Street .  
[ref] This time the fall in stocks on Wall Street is responsible for the drop .

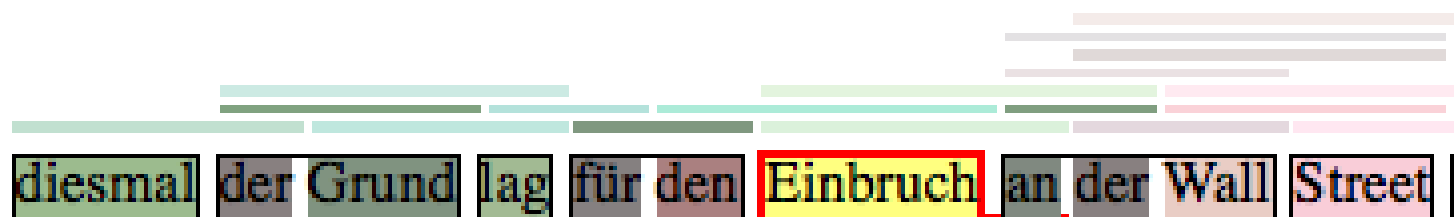
Color highlighting to indicate n-gram overlap with reference translation

darker bleu = word is part of larger n-gram match

## Analysis: Input Annotation

100 occurrences in corpus, 52 distinct translations, translation entropy: 3.08447

[#4]



- For each word and phrase, color coding and stats on
  - number of occurrences in training corpus
  - number of distinct translations in translation model
  - entropy of conditional translation probability distribution  $\phi(e|f)$  (normalized)

# Analysis: Bilingual Concordancer

106



## entre autres(560/1554)

...d and made recommendations , " **inter alia** " , with respect to the follow...  
...on ( EC ) No 1995 / 2000 imposing , **inter alia** , a definitive anti @-@ dumping dut...  
...ervices . this increase , arising , **inter alia** , as a result of economic growth , ...  
...of paragraph 1 the Commission may , **inter alia** , bring forward :  
... of stocks of obsolete pesticides , **inter alia** , by supporting projects aimed at s...  
...wn rules of procedure which shall , **inter alia** , contain provisions for convening ...  
...uch specific agreements may cover , **inter alia** , financing provisions , assignment...  
...he internal market and concerning , **inter alia** , health and environmental protecti...  
...e product concerned ) originating , **inter alia** , in Belarus and Russia ( the count...  
...e product concerned ) originating , **inter alia** , in India .

... des recommandations concernant , **entre autres** , les questions spécifiques suiva...  
...995 / 2000 du Conseil instituant , **entre autres** , un droit antidumping définitif ...  
...nsports . cette augmentation , due **entre autres** facteurs à la croissance économi...  
...aragraphe 1 , la Commission peut , **entre autres** , présenter :  
...r les stocks de vieux pesticides , **entre autres** en soutenant des projets à cet ef...  
...lement intérieur , qui contient , **entre autres** dispositions , les modalités de c...  
...ords spécifiques peuvent porter , **entre autres** , sur les mécanismes financiers s...  
...hé intérieur et qui concernent , **entre autres** , la santé et la protection de l&...  
...it concerné " ) originaire , **entre autres** , du Belarus et de Russie ( ci @-@ ...  
...t concerné " ) originaires , **entre autres** , de l ' Inde .

## notamment(447/1554)

... the EU budget by addressing " **inter alia** " the problems of accountabili...  
...ates , the Commission has adopted , **inter alia** , Decision 2003 / 526 / EC ( 3 ) wh...  
...d equitable development involving , **inter alia** , access to productive resources , ...  
...ertain products which could be used **inter alia** , as equipment on board ships but w...  
...nexes , taking into consideration , **inter alia** , available scientific , technical ...  
...w that it is absolutely necessary , **inter alia** , because of enlargement , to find ...  
...paragraphs 1 and 2 as appropriate , **inter alia** , by conducting studies and compili...  
...liability and efficiency , caused , **inter alia** , by insufficient technical and adm...  
...in the Programme shall be pursued , **inter alia** , by the following means :

...get de l' Union , ce qui passe **notamment** par la résolution du problème de r...  
...es États membres , la Commission a **notamment** arrêté la décision 2003 / 526 / C...  
... durable et équitable , impliquant **notamment** l' accès aux ressources produc...  
...usceptibles d' être utilisés **notamment** comme équipements mis à bord , mai...  
...ion et à ses annexes , compte tenu **notamment** des informations scientifiques , tec...  
...os ; il est absolument nécessaire , **notamment** en raison de l' élargissement ...  
...ragraphes 1 et 2 le cas échéant , **notamment** en menant des études et en compilan...  
... et d' efficacité en raison , **notamment** , d' une interopérabilité tec...  
...nis dans le programme , il convient **notamment** de mettre en oeuvre les moyens ci @-...

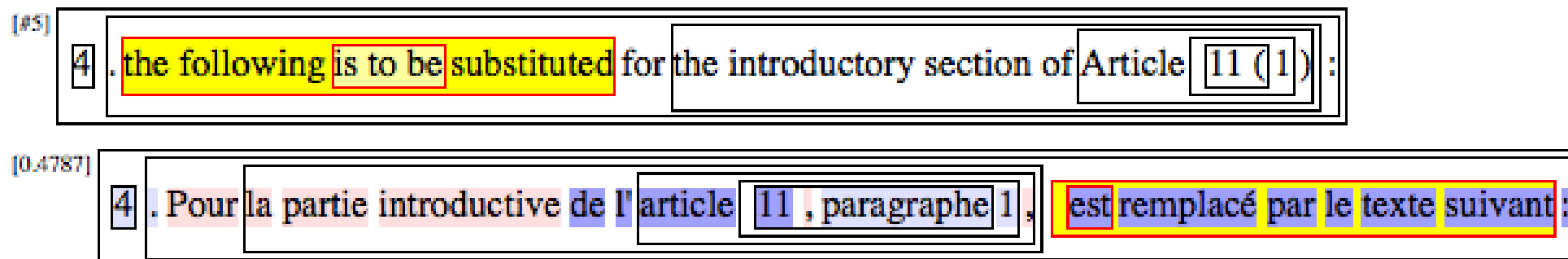
translation of input phrase in training data context

## Analysis: Alignment

diesmal	der Grund	lag	für den	Einbruch	an der	Wall	Street	.
52] This time	was	the reason	for the	collapse	on	Wall	Street	.

Phrase alignment of the decoding process  
(red border, interactive)

# Analysis: Tree Alignment



Uses nested boxes to indicate tree structure  
(red border, yellow shaded spans in focus, interactive)  
for syntax model, non-terminals are also shown

# Analysis: Comparison of 2 Runs

## annotated sentences

sorted by [order](#) order [worse](#) display [fullscreen](#) showing 5 [more](#) [all](#)

identical same better worse

2348 51 57 69

93% 2% 2% 3%

[2143:0.2974] In Austria , Haider and Co. are ready to govern to prevent a red and black coalition .

[2143:0.1754] In Austria , Haider and Co. are prepared to rule to prevent a red and black coalition .

[ref] Haider and his party are ready to govern Austria in order to avoid red @-@ black coalition .

---

[2165:0.3174] The SPÖ wants to show that the cooperation of both parties is possible - in some countries and in the social partnership that is already the case .

[2165:0.2061] The SPÖ wants to show that a cooperation of both parties is possible - in some countries and in the social partnership that is already the case .

[ref] SPÖ would like to show that the cooperation of the two parties is possible - it does exist in some of the provinces as well as in social partnership .

Different words are highlighted  
sortable by most improvement, deterioration

# Acknowledgements

110



# Moses Developers

Abhishek Arun	Adam Lopez	Ales Tamchyna	Alex
Amittai Axelrod	Ankit Srivastava	Anthony Rousseau	Benjamin Gottesman
Barry Haddow	Ondrej Bojar	Chris Callison-Burch	Christine Corbett
Christian Hardmeier	Christian Federmann	Lane Schwartz	David Talbot
Edmund Huber	Evan Herbst	Andreas Eisele	Eva Hasler
Frederic Blain	Brooke Cowan	Grace M. Ngai	Kenneth Heafield
Hieu Hoang	H. Leal Fontes	Holger Schwenk	Josh Schroeder
Jean-Baptiste Fouet	Joern Wuebker	Jorge Civera	Konrad Rawlik
Abby Levenberg	Alexandra Birch	Bo Fu	M.J.Bellino-Machado
Mauro Cettolo	Marcello Federico	Michael Auli	John Joseph Morgan
Mark Fishel	Gabriele Antonio Musillo	Miles Osborne	Nadi Tomeh
Nicola Bertoldi	Oliver Wilson	Pascual Martinez	Philipp Koehn
Phil Williams	Bruno Pouliquen	Raphael Payen	Chris Dyer
Joao Lus Rosas	Rico Sennrich	Herve Saint-Amand	Felipe Sanchez Martinez
Sara Stymne	Steven B. Parks	Steven Buraje Poggel	Andre Lynum
Yizhao Ni	David Kolovratnak	Sergio Penkale	Stephan
Suzy Howlett	Wade Shen	Yang Gao	Tsuyoshi Okita
Alexander Fraser	Richard Zens		