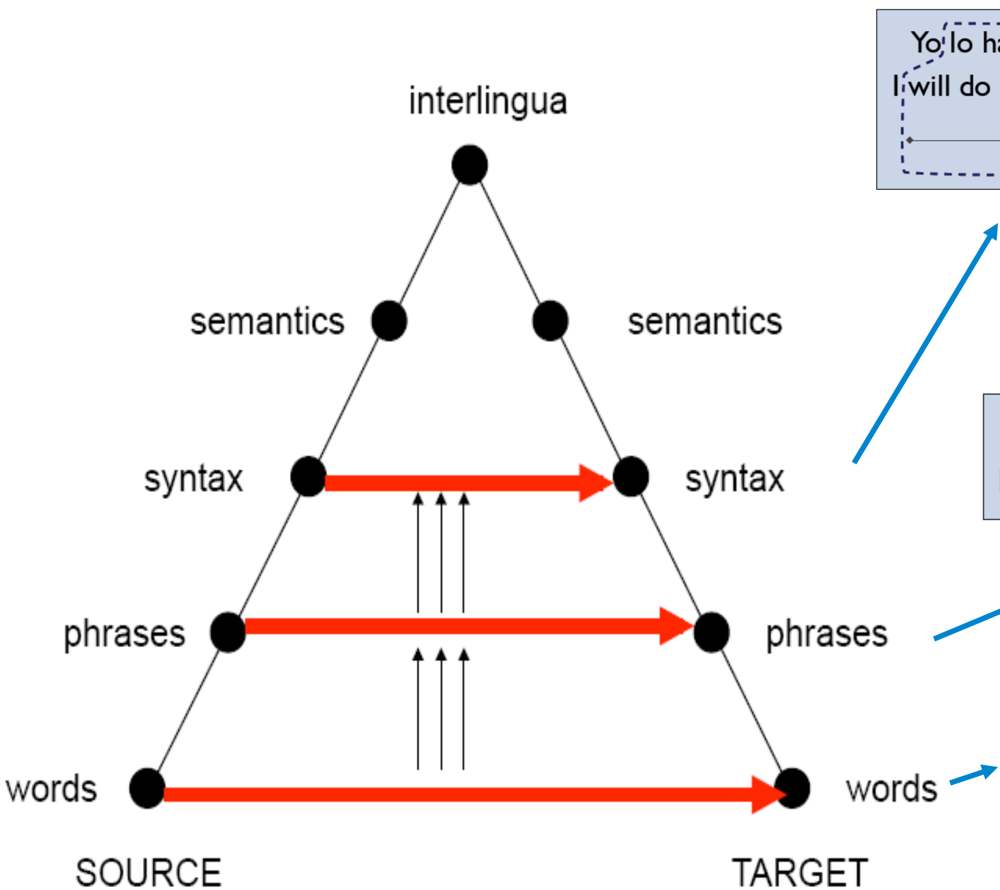
# CSE 517
# Natural Language Processing
# Winter 2013

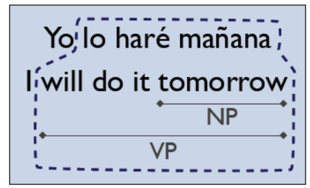
## Syntax-Based Translation

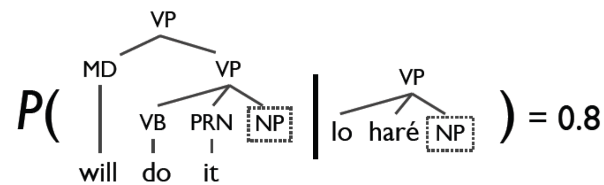
Luke Zettlemoyer

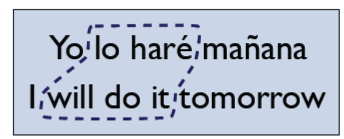
Slides from Philipp Koehn, Matt Post

# Levels of Transfer

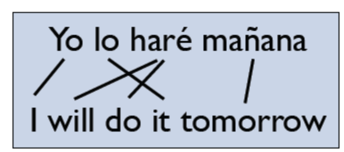



| *English* (E) | *P*( E \| lo haré ) |
|---|---|
| will do it | 0.8 |
| will do so | 0.2 |



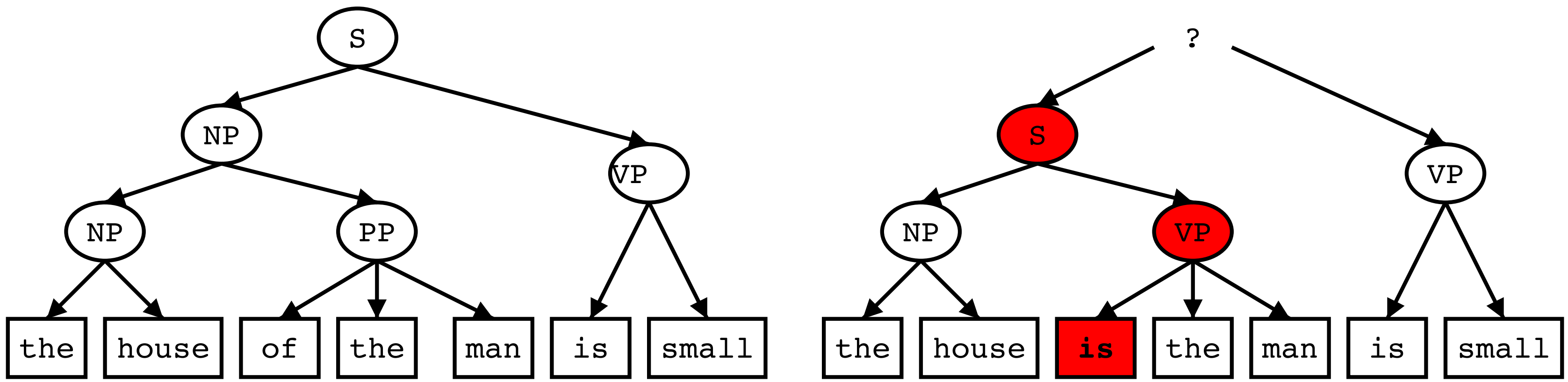

| *English* (E) | *P*( E \| mañana ) |
|---|---|
| tomorrow | 0.7 |
| morning | 0.3 |

# Goals of Translating with Syntax

- Reordering driven by syntactic
  - E.g., move German verb to final position
- Better explanation for function words
  - E.g., prepositions and determiners
- Allow long distance dependencies
  - Translation of verb may depend on subject or object, which can have high string distance
- Will allow for the use of syntactic language models

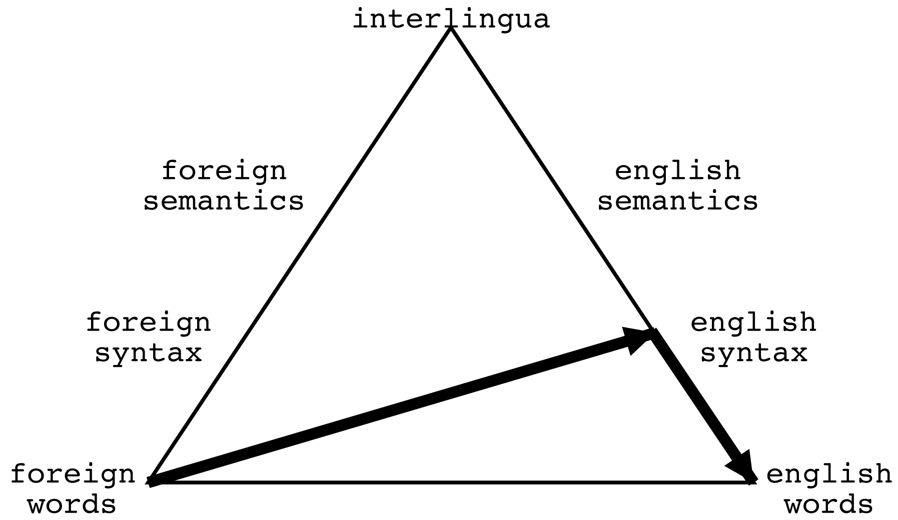
# Syntactic Language Models

- Allows for long distance dependencies




- Left translation would be preferred!

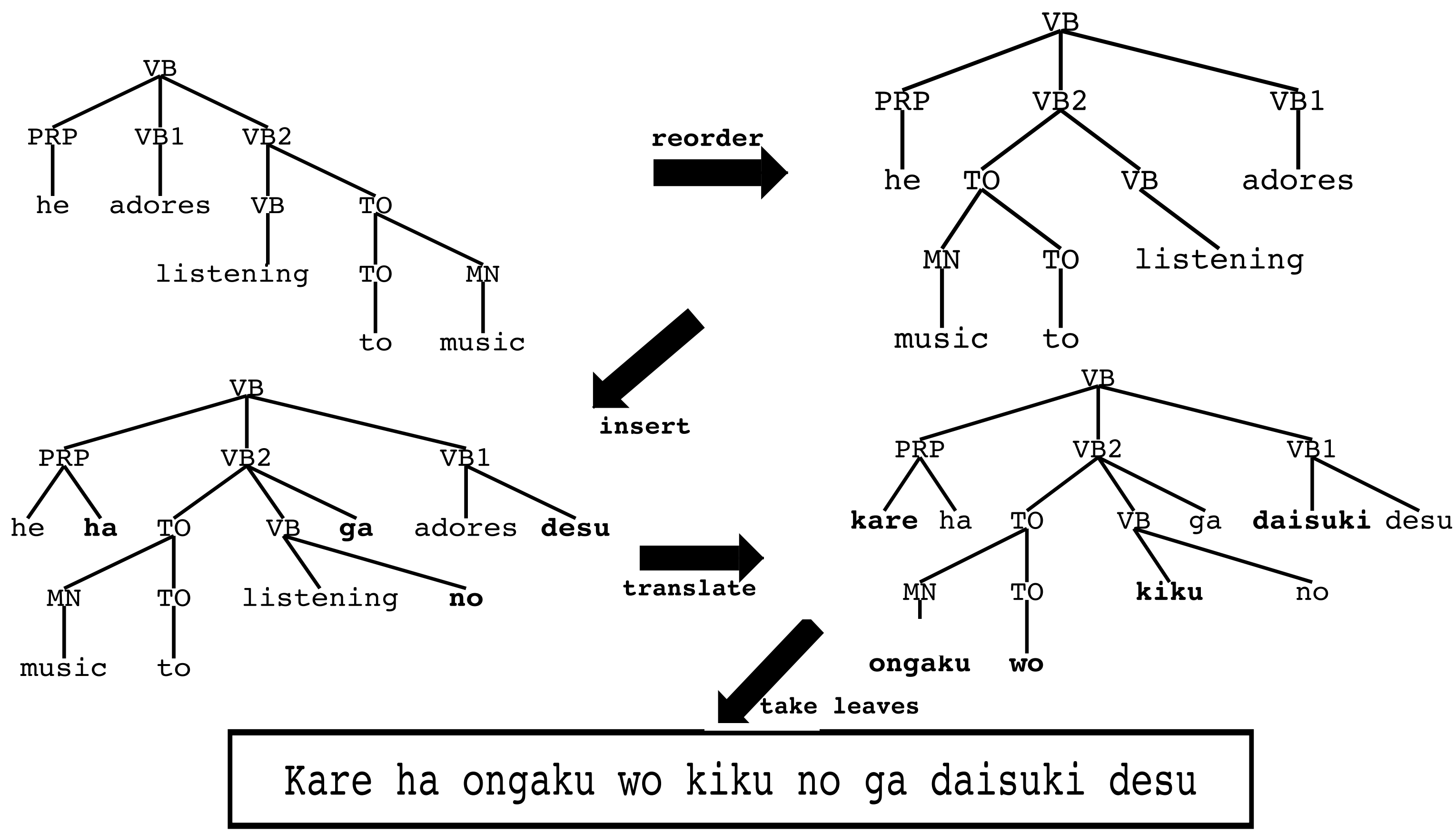
# String to Tree Translation




- Create English syntax trees during translation [Yamada and Knight, 2001]
  - very early attempt to learn syntactic translation models
  - use state-of-the-art parsers for training
  - allows us to model translation as a parsing problem, reusing algorithms, etc.

# Yamada and Knight [2001]

- p(f|e) is a generative process from an English tree to a foreign string



# Learned Model

- Reordering Table

| Original Order | Reordering | p(reorder\|original) |
|---|---|---|
| PRP VB1 VB2 | PRP VB1 VB2 | 0.074 |
| **PRP VB1 VB2** | **PRP VB2 VB1** | **0.723** |
| PRP VB1 VB2 | VB1 PRP VB2 | 0.061 |
| PRP VB1 VB2 | VB1 VB2 PRP | 0.037 |
| PRP VB1 VB2 | VB2 PRP VB1 | 0.083 |
| PRP VB1 VB2 | VB2 VB1 PRP | 0.021 |
| VB TO | VB TO | 0.107 |
| **VB TO** | **TO VB** | **0.893** |
| TO NN | TO NN | 0.251 |
| **TO NN** | **NN TO** | **0.749** |

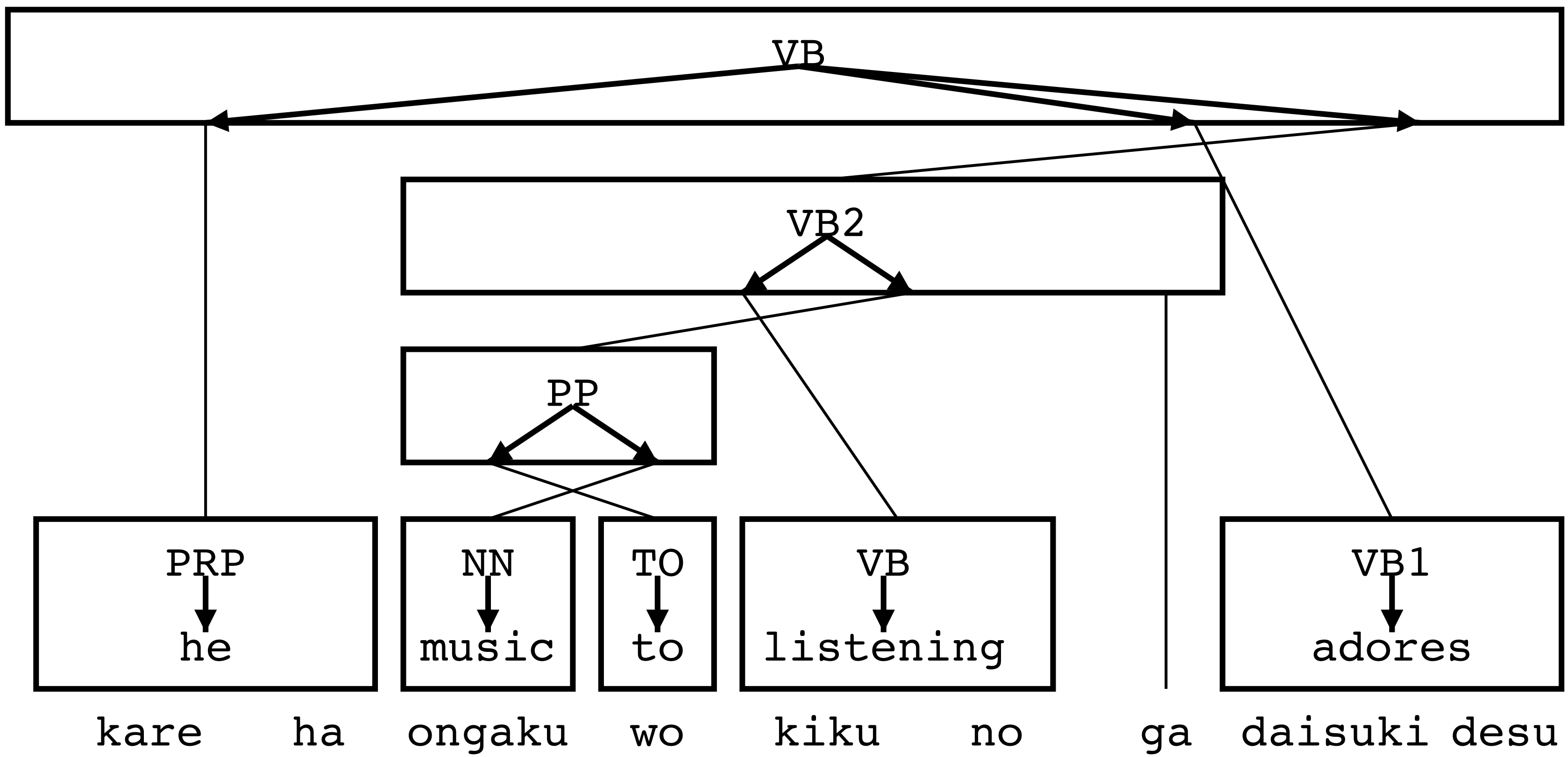
# Yamada and Knight: Decoding

- A Parsing Problem
  - Can use CKY Algorithm, with rules that encode reordering, inserted works

kare ha ongaku wo kiku no ga daisuki desu

# Yamada and Knight: Decoding

- A Parsing Problem
  - Can use CKY Algorithm, with rules that encode reordering, inserted works



# Yamada and Knight: Training

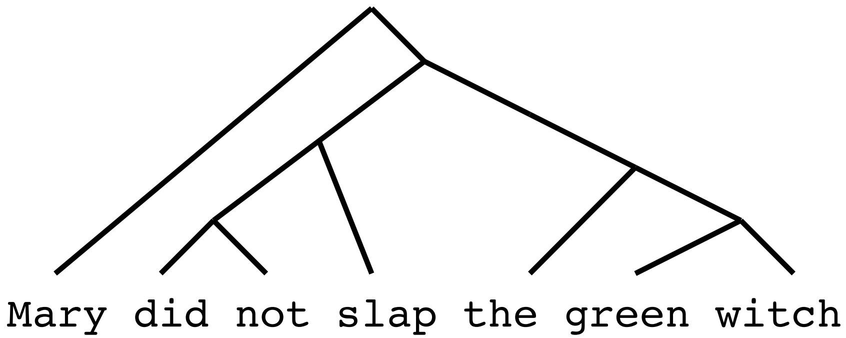
- Want P(f|e), where e is a English parse tree
  - Parse the English side of bi-text
  - Use parser output as gold standard
- Many different derivations from e to f (for a fixed pair)
  - Use EM training approach
  - Same idea as IBM Models (but a bit more complex)

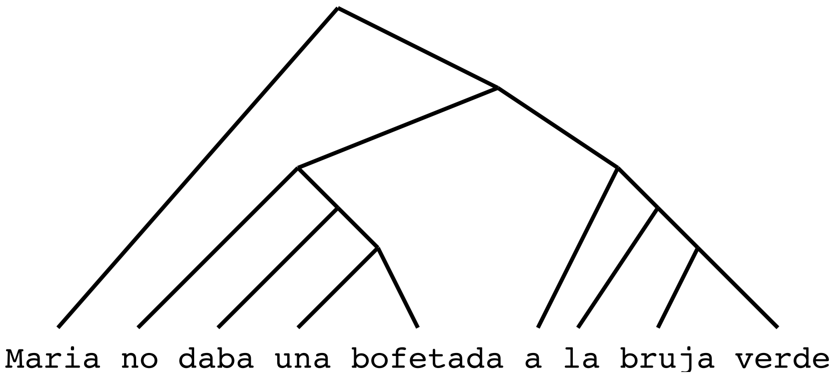
# Is The Model Realistic?

- Do English trees align well onto foreign string?
- Crossings between French-English [Fox, 2002]
  - ~1-5 per sentence (depending on how you count)
- Can be reduced by
  - Flattening tree, as done by Yamada and Knight
  - Mixing in phrase level translations
  - Special casing many constructions

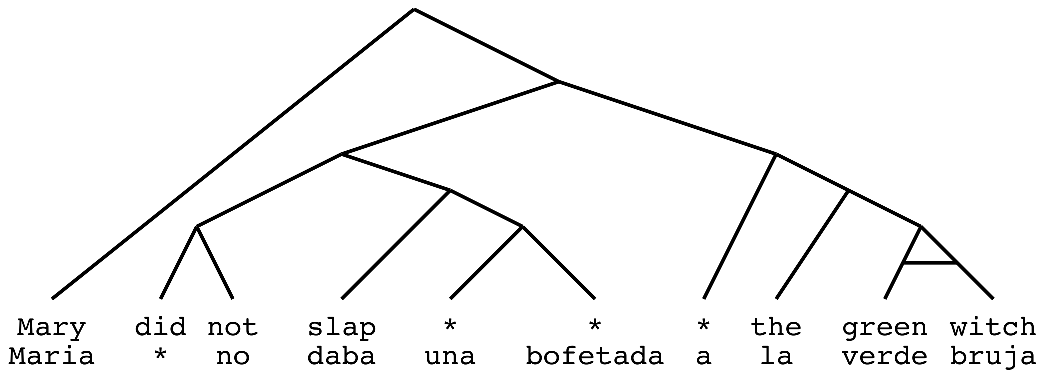
# What about tree-to-tree?

- Consider the following trees:



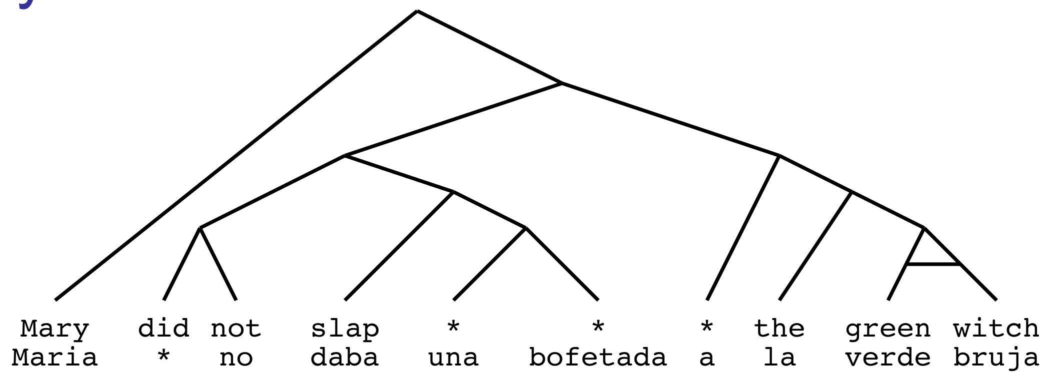




- We might merge them as follows:



# Inversion Transduction Grammars (ITGs)

- Simultaneously generates two trees (English and Foreign) [Wu, 1997]
- Rules, binary and unary
  - $X \rightarrow X_1X_2 \mid\mid X_1X_2$
  - $X \rightarrow X_1X_2 \mid\mid X_2X_1$
  - $X \rightarrow e \mid\mid f$
  - $X \rightarrow e \mid\mid *$
  - $X \rightarrow * \mid\mid f$



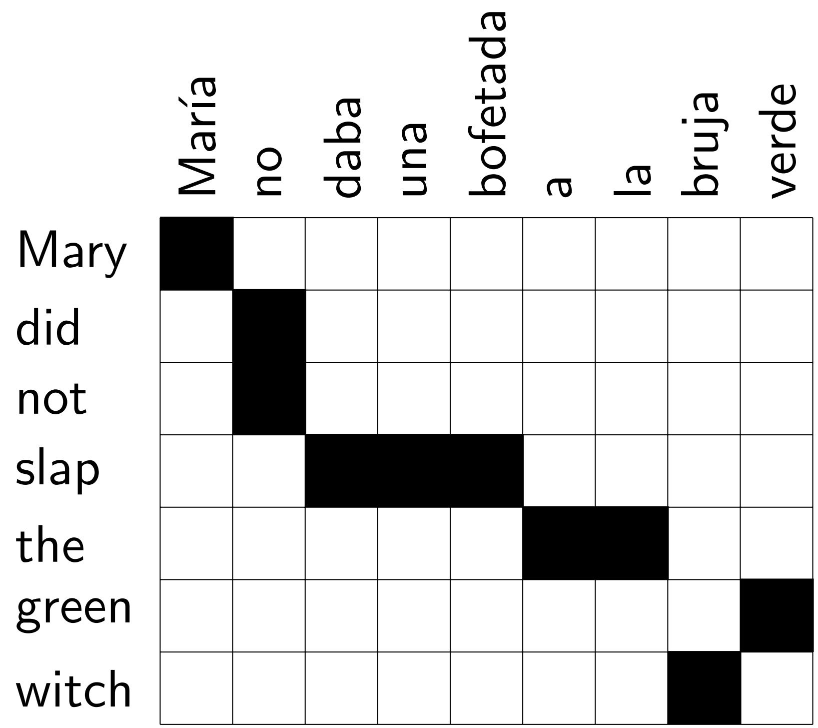

- Builds a common binary tree
  - Limits the possible reorderings
  - Challenging to model complete phrases
  - But, can do decoding as parsing, just like before!

# Hierarchical Phrase Model [Chiang, 2005]

- Hybrid of ITGs and phrase based translation
- Word rules
  - X → maison || house
- Phrasal Rules
  - X → daba una bofetada || slap
- Mixed Terminal / Non-terminal Rules
  - X → X bleue || blue X
  - X → ne X pas || not X
  - X → $X_1$ $X_2$ || $X_2$ of $X_1$
- Technical Rules
  - S → S X || S X
  - S → X || X

# Hierarchical Rule Extraction

- Include all word and phrase alignments
  - X→ verde || green
  - X→ bruja verde || green witch
  - …
- Consider every possible rule, with variable for subphrases
  - X → X verde || green X
  - X → bruja X || X witch
  - X → a la X || the X
  - X → daba una botefada || slap X
  - …

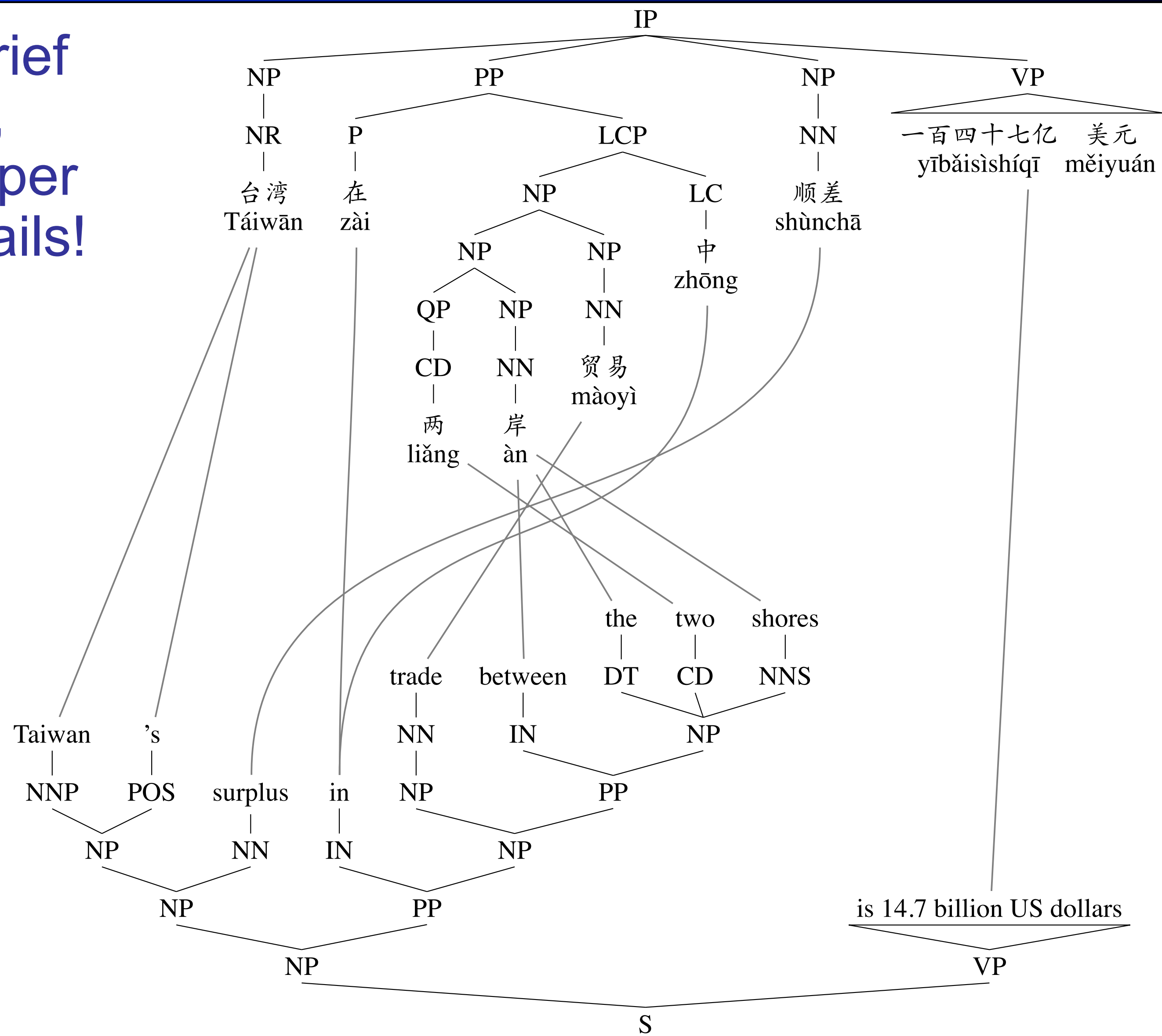


# The Rest of The Details

- See paper [Chiang, 2005]
  - Model is done much like phrase-based systems
  - Too many rules → Need to prune
- Efficient parsing algorithms for decoding
- How well does it work?
  - Chinese-English: 26.8 → 28.8 BLEU
  - Competitive with phrase-based systems on most other language pairs, but lags behind when the language pair has modest reordering
  - There has been significant work on better ways of extracting translation rules, and estimating parameters

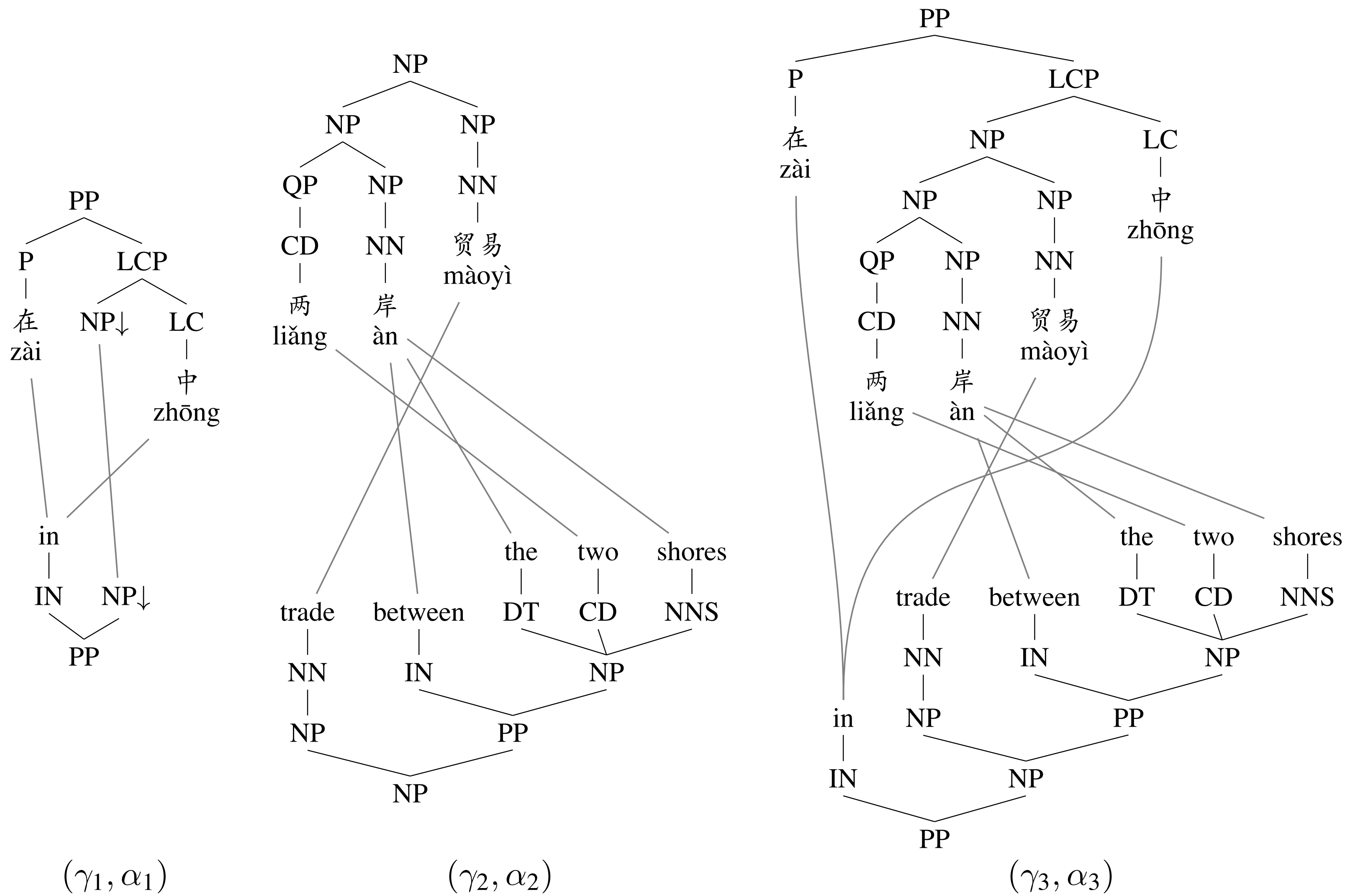
# Tree to Tree Translation [Chiang, 2010]

- Very brief sketch, see paper for details!



# Tree to Tree Translation

- Key idea: Learn synchronous tree substitution grammar



# Tree to Tree Translation

- To make it work: Allow many different tree structures (when syntax doesn't align directly)

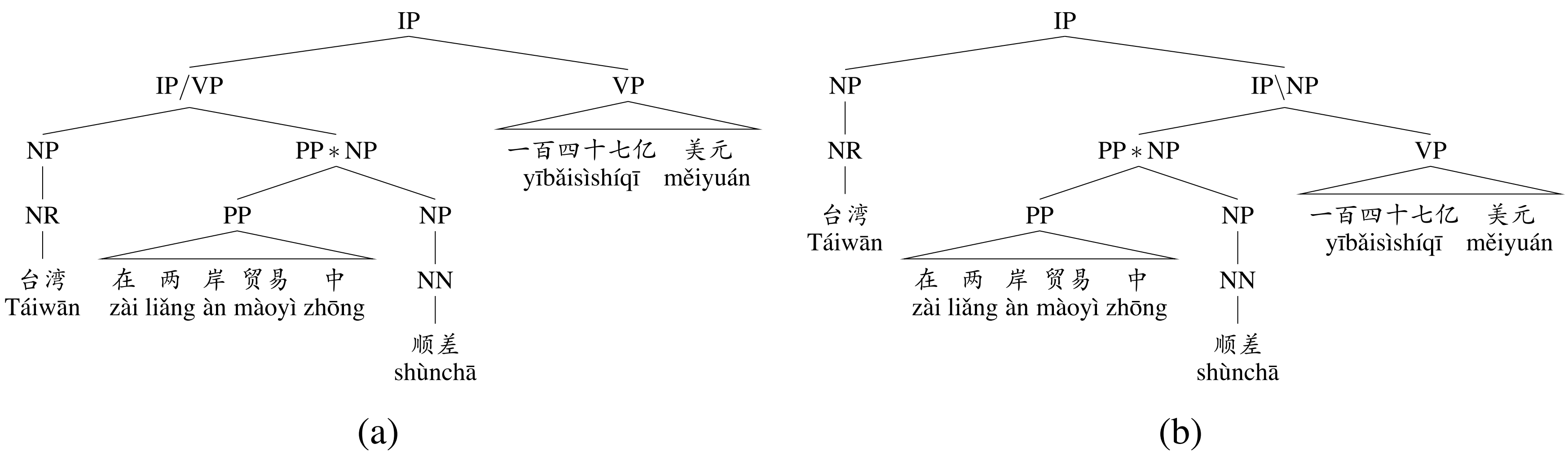



Figure 4: Fuzzy tree-to-tree extraction effectively restructures the Chinese tree from Figure 2 in two ways but does not commit to either one.

# Tree to Tree Translation

- And, the paper has tons of other details…
- But, lets see the results!

| task | extraction | dist. lim. | rules | features | BLEU dev | BLEU test |
|---|---|---|---|---|---|---|
| Chi-Eng | string-to-string | 10 | 440M | 1k | 32.7 | 23.4 |
| | string-to-string | 20 | 440M | 1k | 33.3 | 23.7 |
| | tree-to-tree exact | 20 | 50M | 5k | 32.8 | 23.9 |
| | tree-to-tree fuzzy | 20 | 440M | 160k | 33.9 | 24.3 |
| | + nesting | 20 | 180M | 79k | 33.9 | 24.3 |
| Ara-Eng | string-to-string | 10 | 790M | 1k | 48.7 | 48.9 |
| | tree-to-tree exact | 10 | 38M | 5k | 46.6 | 47.5 |
| | tree-to-tree fuzzy | 10 | 790M | 130k | 49.4 | 49.7 |
| | + nesting | 10 | 190M | 66k | 49.2 | 49.8 |

Table 3: On both the Chinese-English and Arabic-English translation tasks, fuzzy tree-to-tree extraction outperforms exact tree-to-tree extraction and string-to-string extraction. Brackets indicate statistically insignificant differences ($p \geq 0.05$).

# Clause Level Restructing

- Approach:
  - Still use phrase-based system
  - First, parse the input sentence and reorder it
  - Then, pass it to the phrase-based translator
- Why?
  - Most long distance re-ordering is at the clause level
  - E.g., English: SVO, Arabic: VSO, German: relatively free order
  - Most other phenomena can be captured by the large phrase tables!

# [Collins, Koehn, and Kucerova, 2005]

Phrase-based models have an overly simplistic way of handling different word orders.

We can describe the linguistic differences between different languages.

Collins defines a set of 6 simple, linguistically motivated rules, and demonstrates that they result in significant translation improvements.

# Pre-ordering Model

Step 1: Reorder the source language

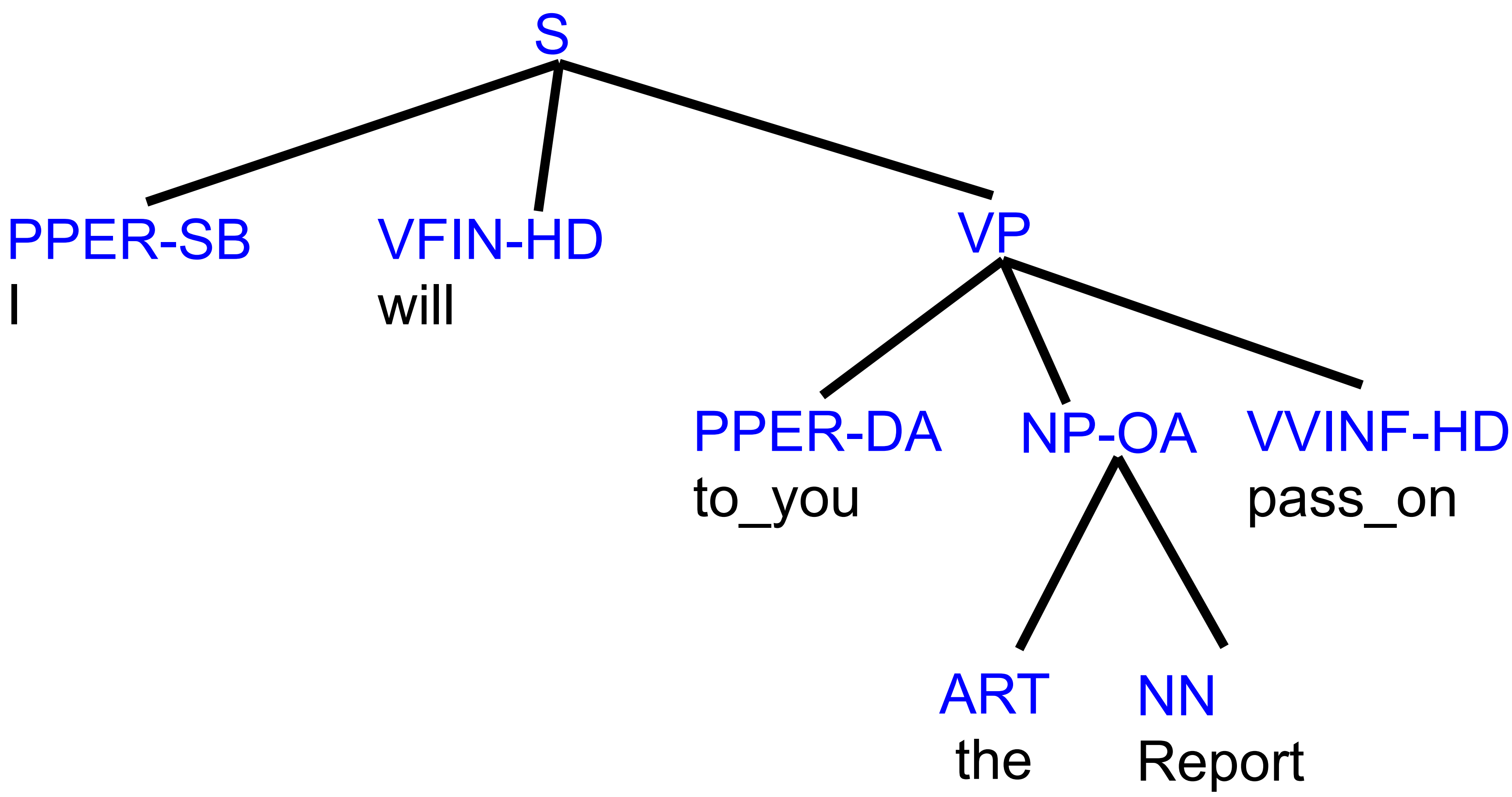
Ich werde Ihnen den Report aushaendigen , damit Sie den eventuell uebernehmen koennen .

Ich werde aushaendigen Ihnen den Report , damit Sie koennen uebernehmen den eventuell .

(I will pass_on to_you the report, so_that you can adopt it perhaps .)

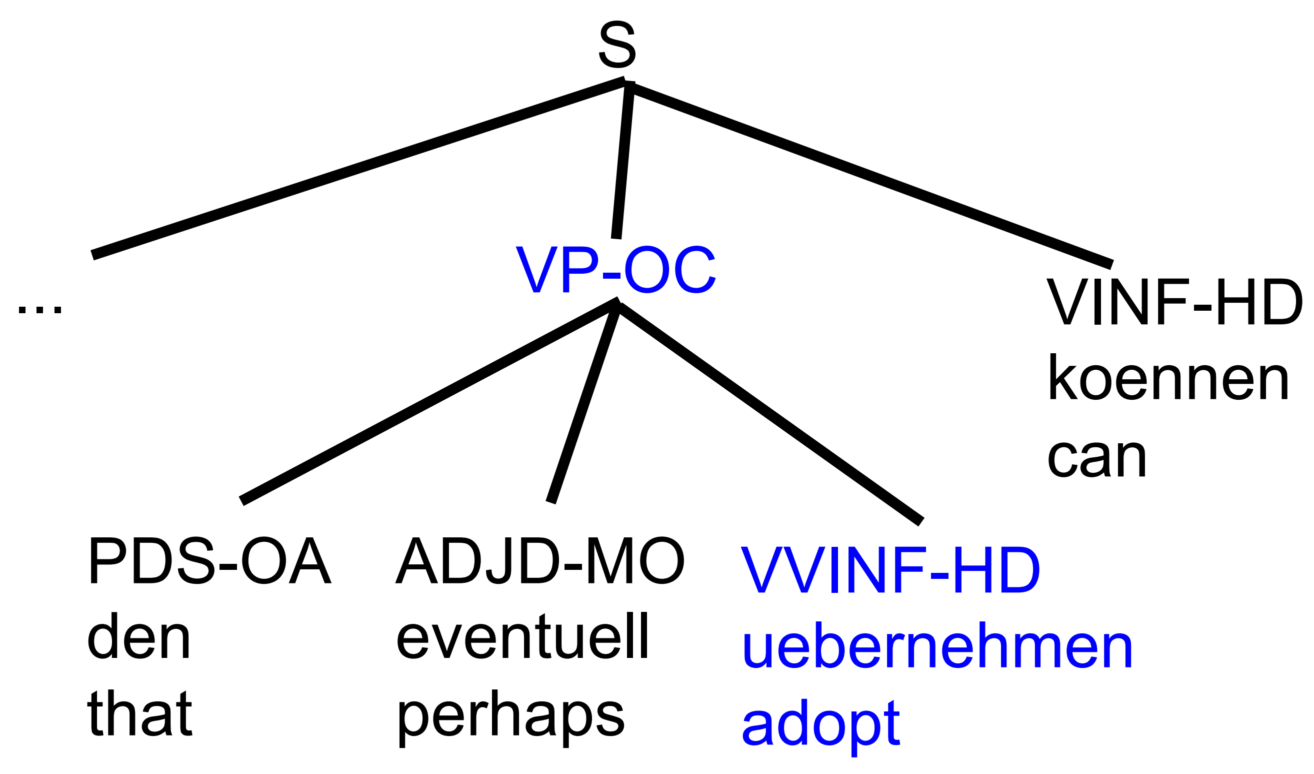
Step 2: Apply the phrase-based machine translation pipeline to the reordered input.

# Example Parse Tree



# Clause Restructuring

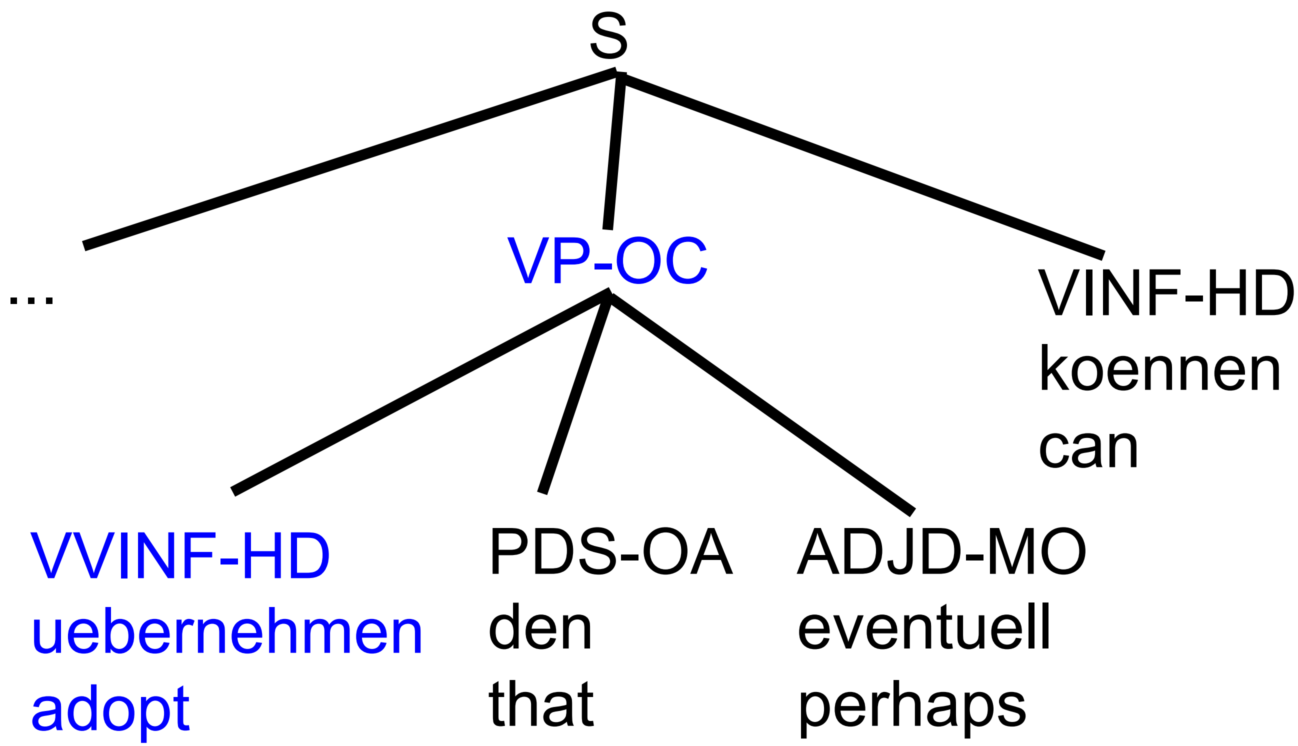
Rule 1: Verbs are initial in VPs
Within a VP, move the head to the initial position



# Clause Restructuring

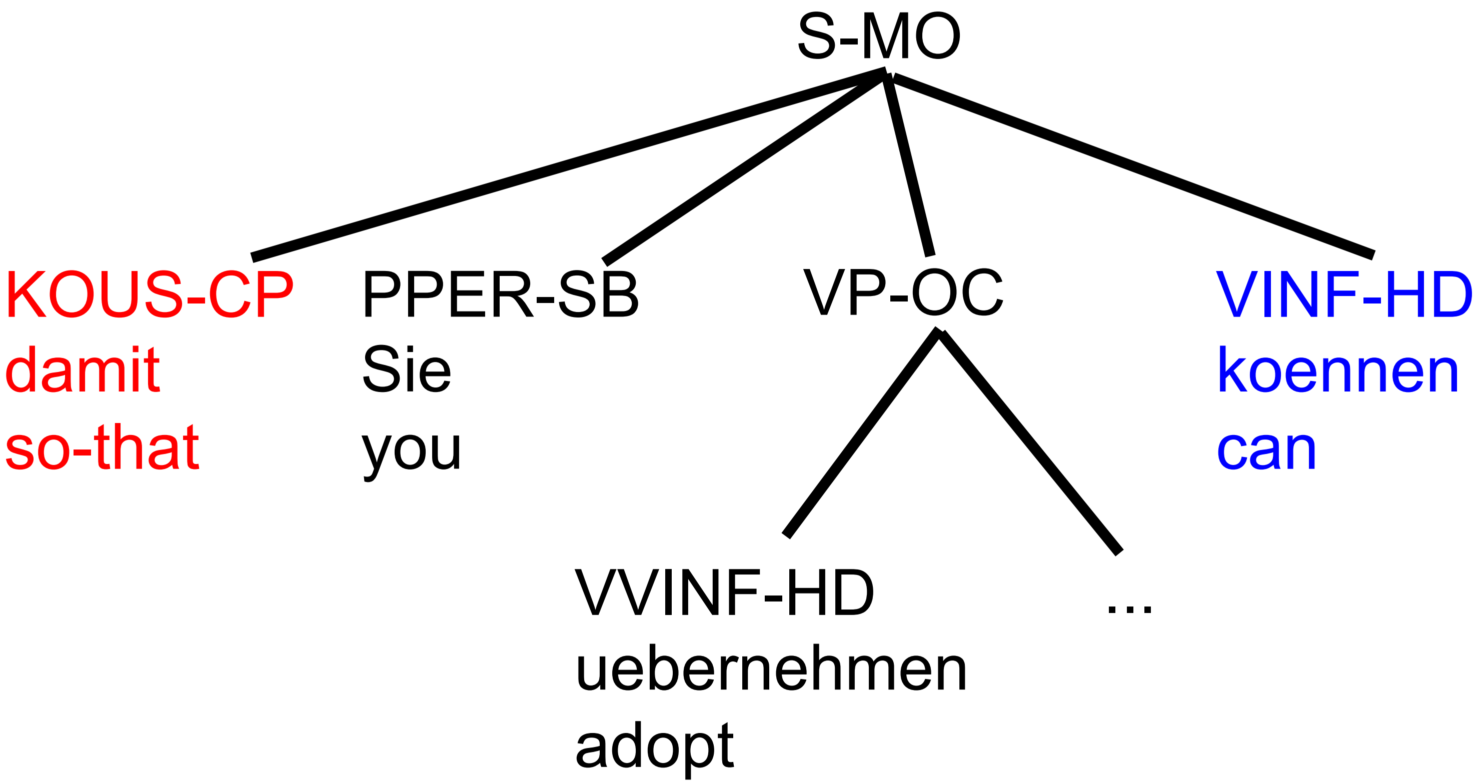
Rule 1: Verbs are initial in VPs

Within a VP, move the head to the initial position



# Clause Restructuring

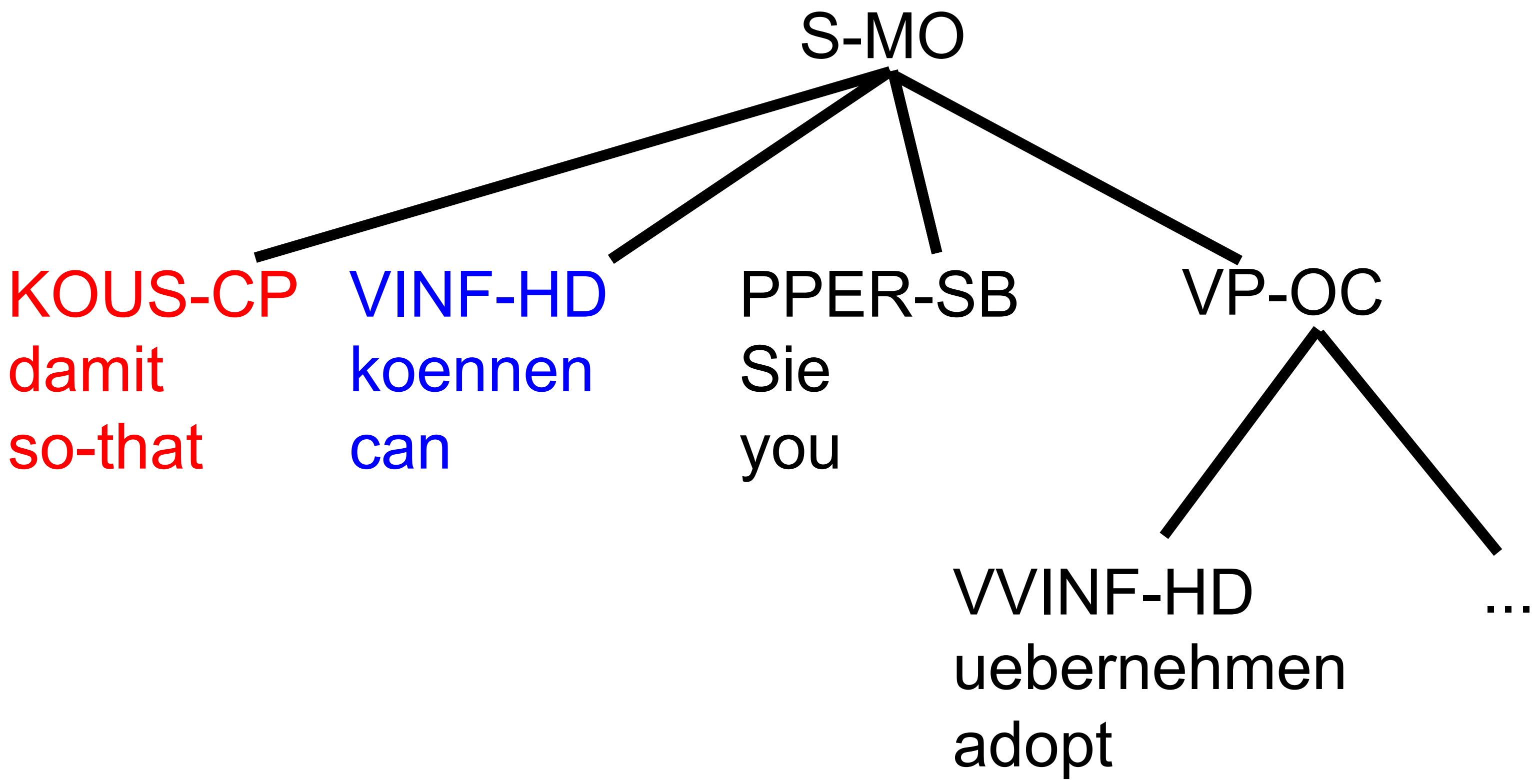
Rule 2: Verbs follow complementizers
In a subordinated clause move the head of the clause to follow the complementizer



# Clause Restructuring

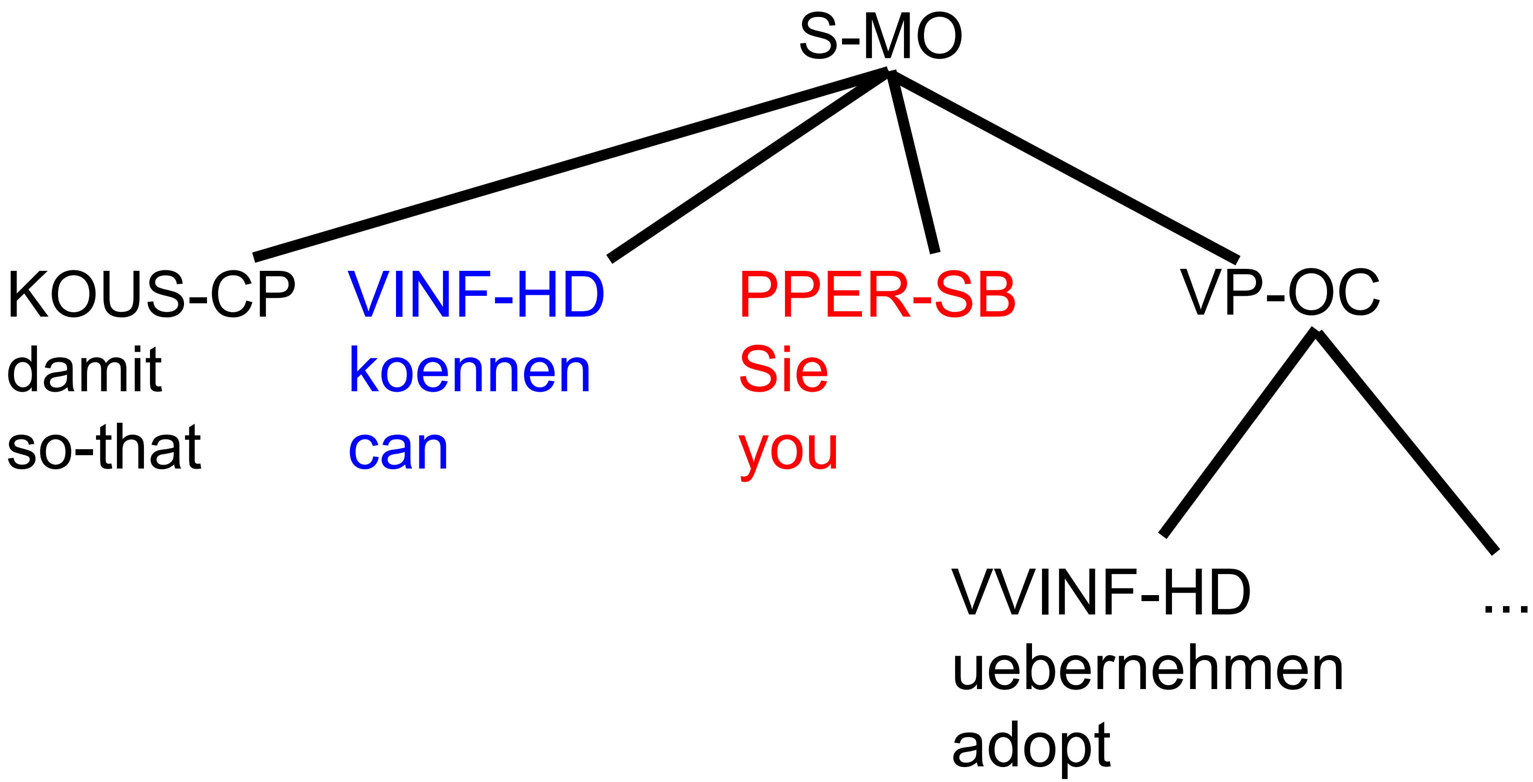
Rule 2: Verbs follow complementizers

In a subordinated clause mote the head of the clause to follow the complementizer



# Clause Restructuring

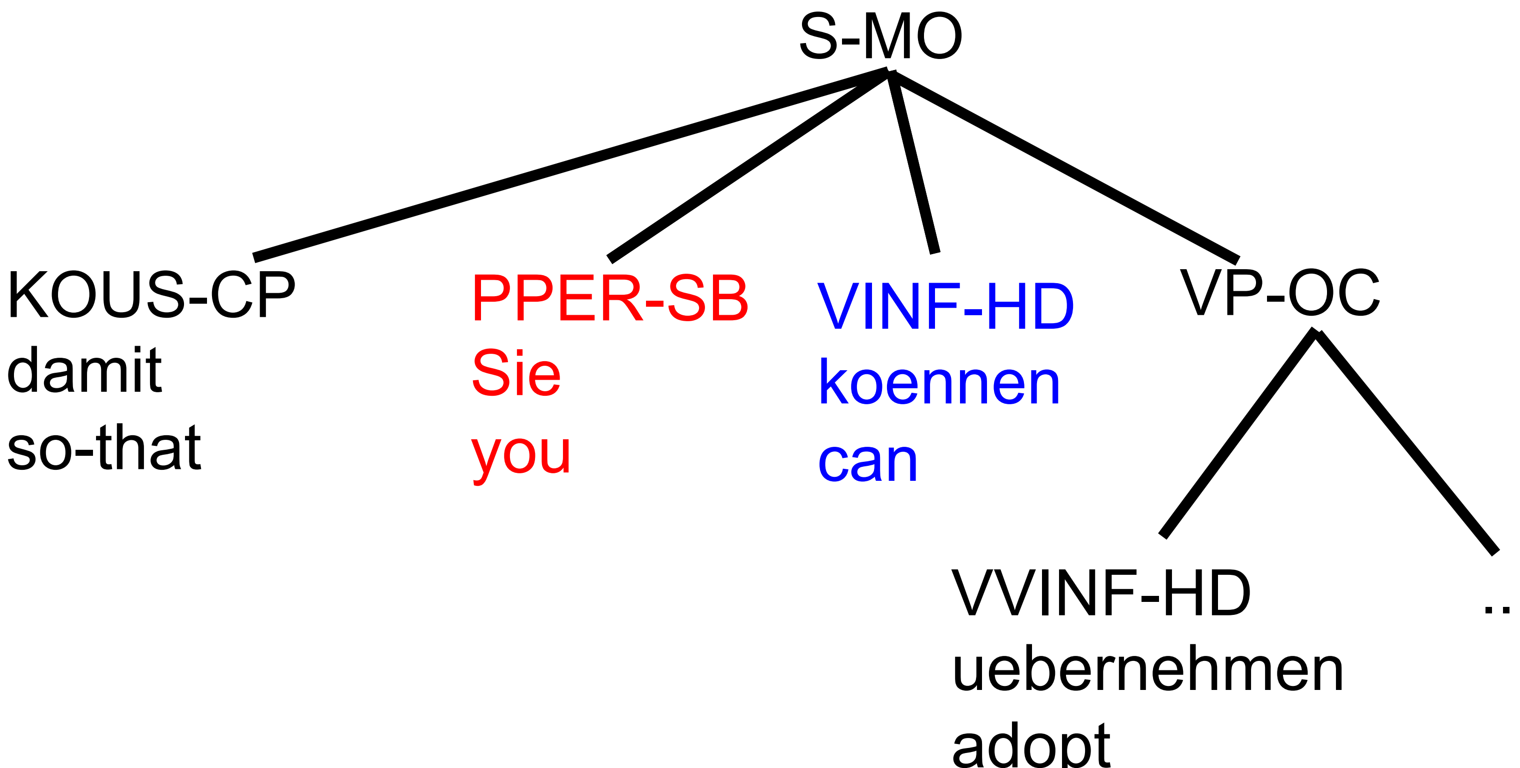
Rule 3: Move subject
The subject is moved to directly precede the head of the clause



# Clause Restructuring

Rule 3: Move subject

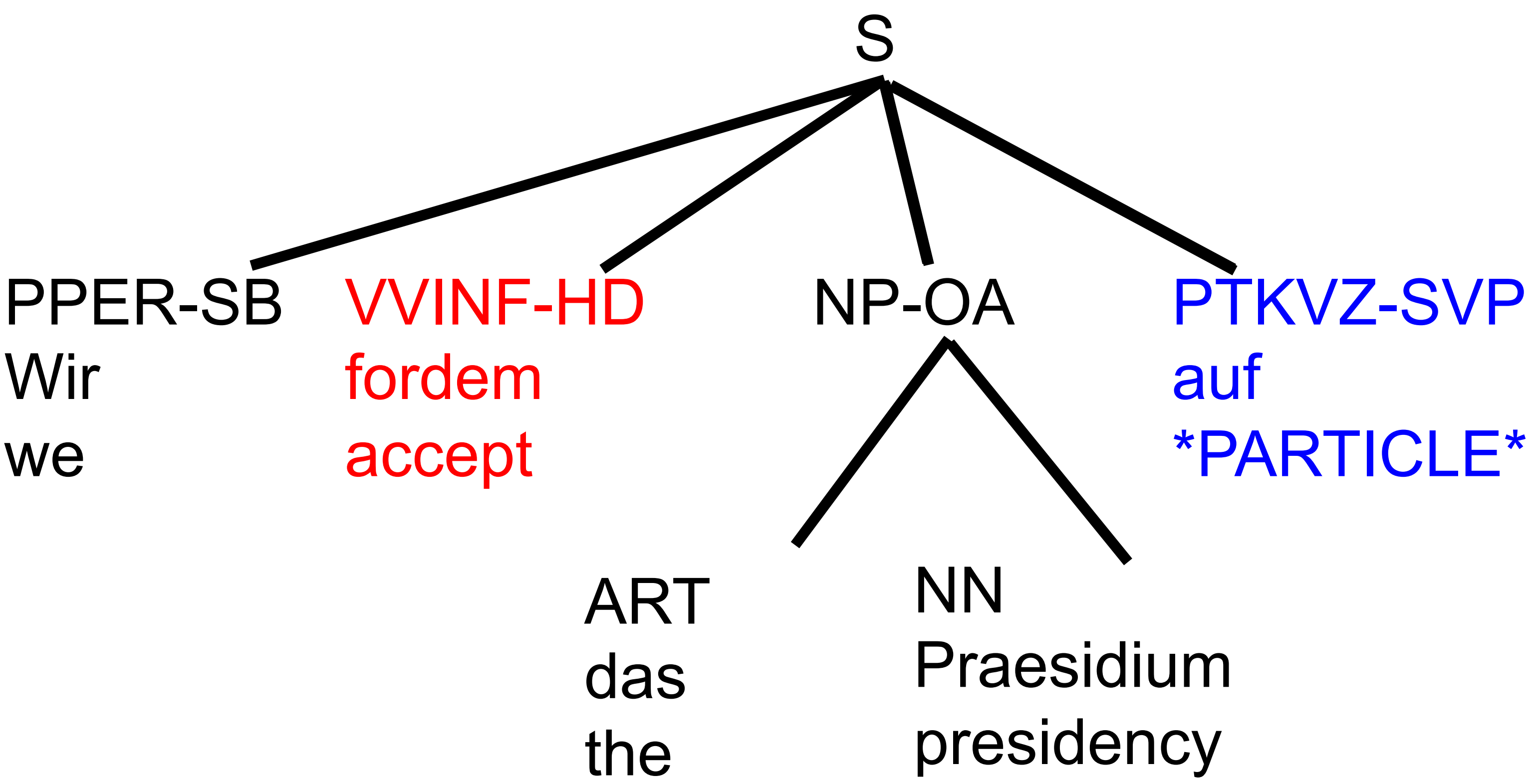
The subject is moved to directly precede the head of the clause



# Clause Restructuring

Rule 4: Particles

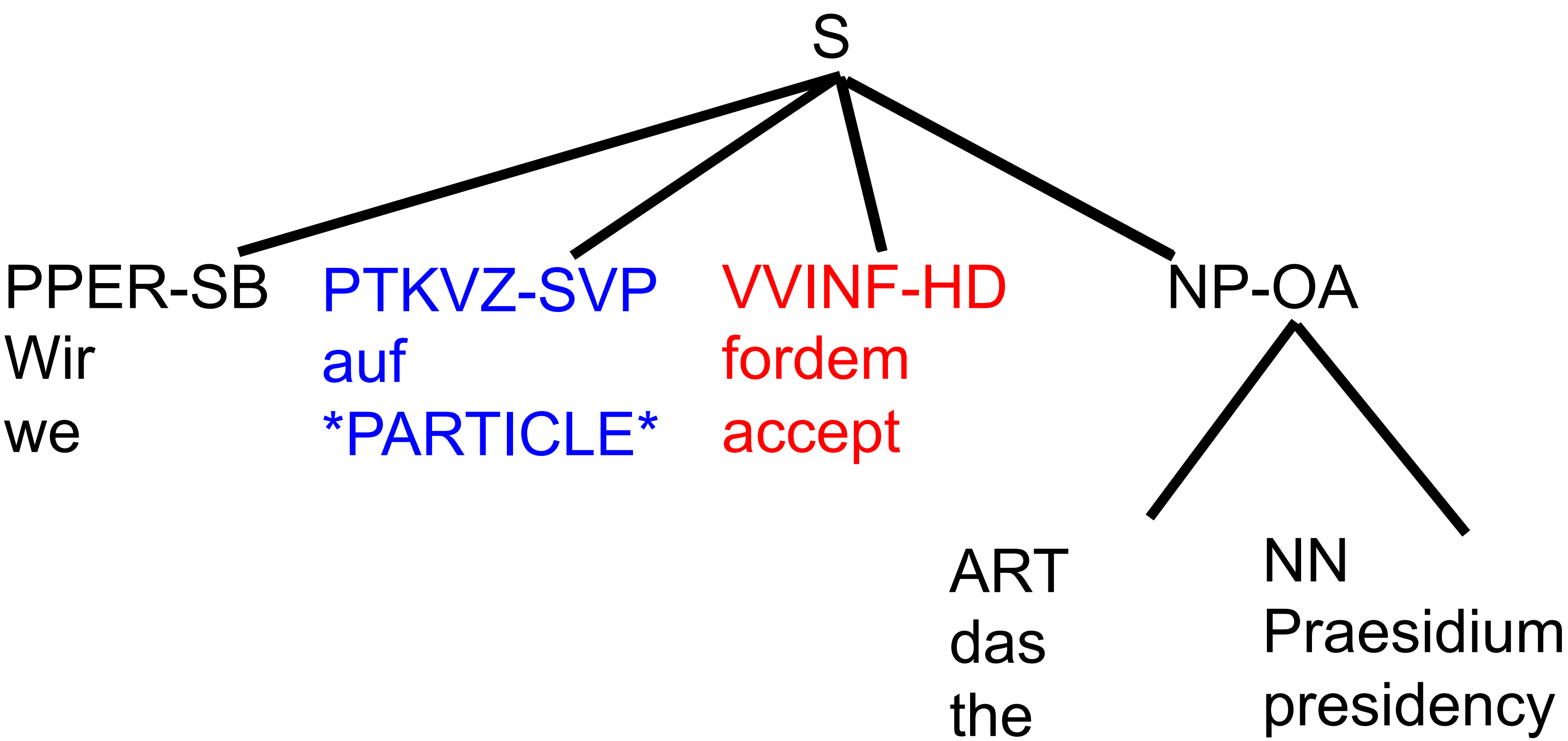
In verb particle constructions, the particle is moved to precede the finite verb



# Clause Restructuring

Rule 4: Particles

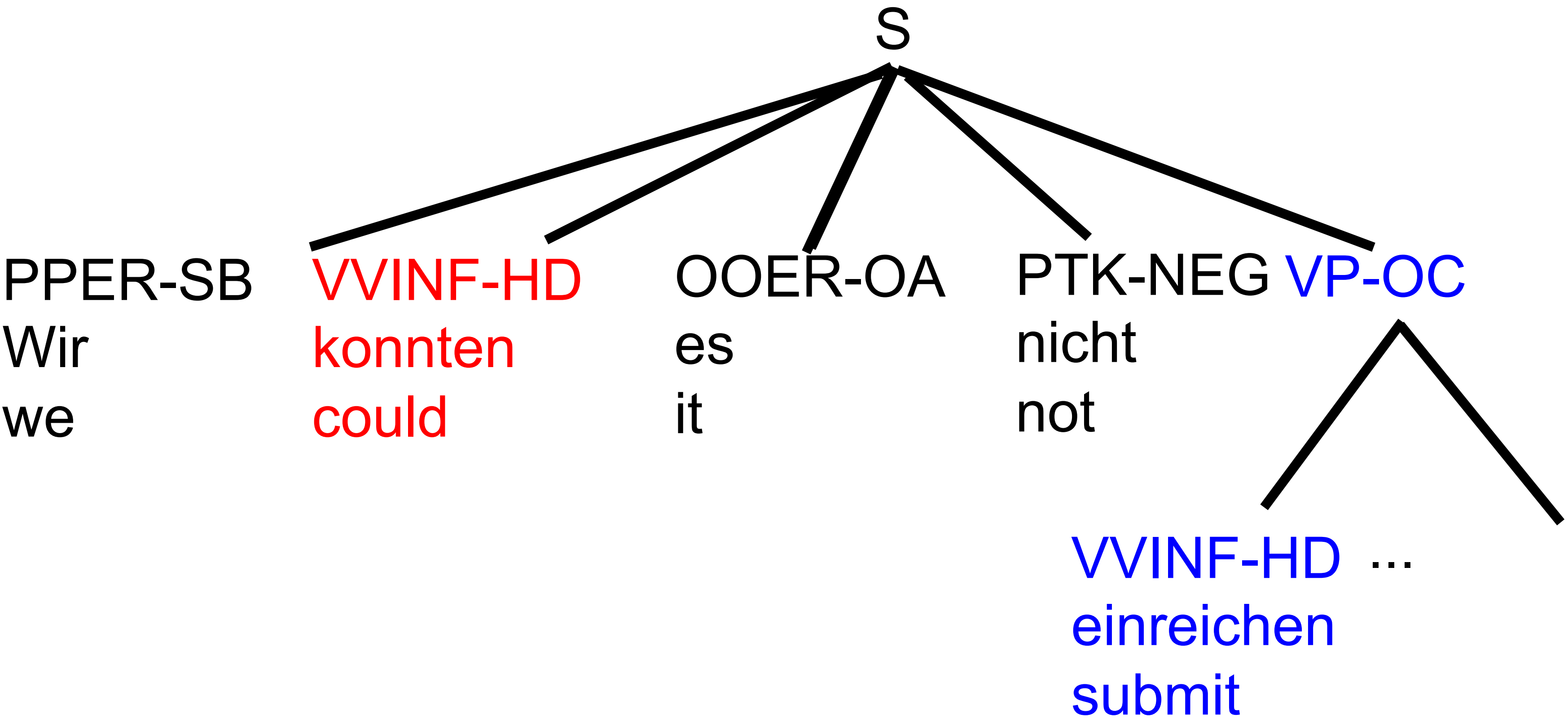
In verb particle constructions, the particle is moved to precede the finite verb



# Clause Restructuring

Rule 5: Infinitives

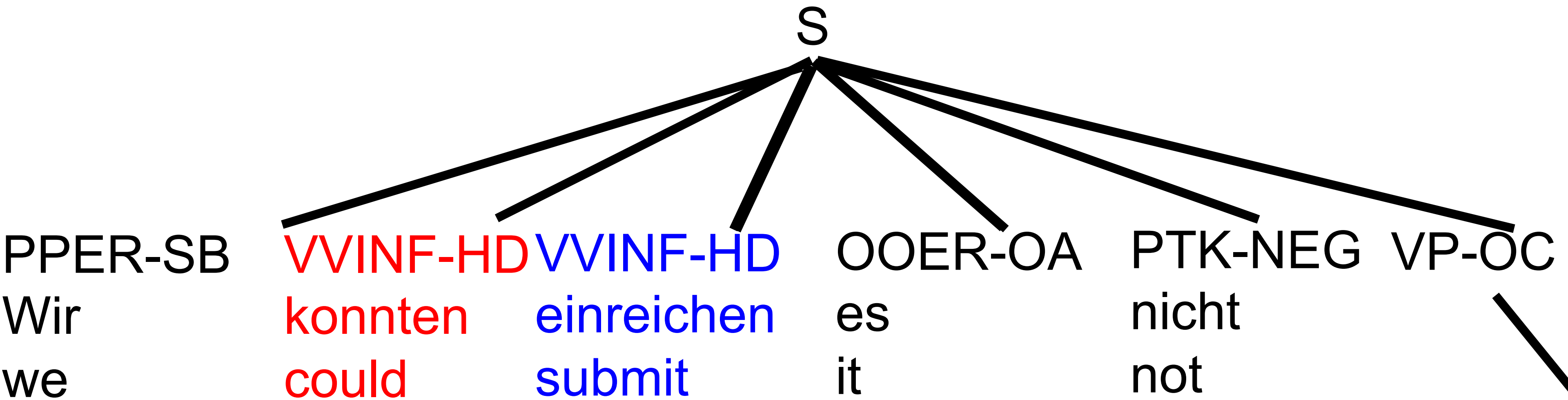
Infinitives are moved to directly follow the finite verb within a clause



# Clause Restructuring

Rule 5: Infinitives

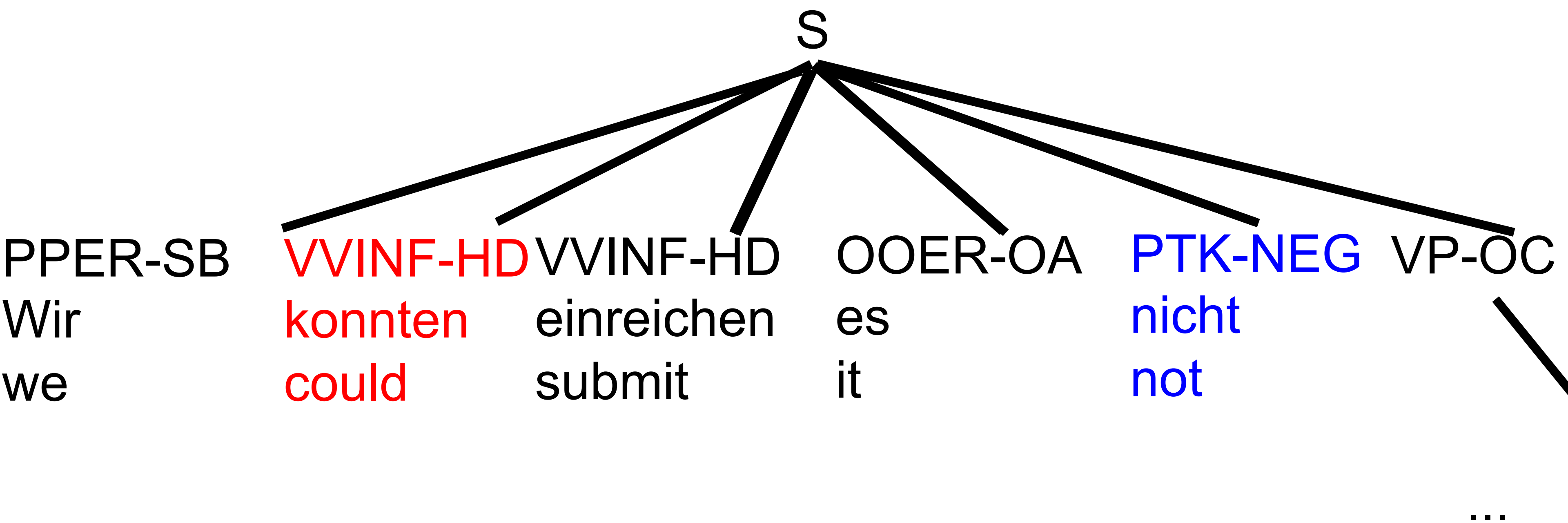
Infinitives are moved to directly follow the finite verb within a clause



# Clause Restructuring

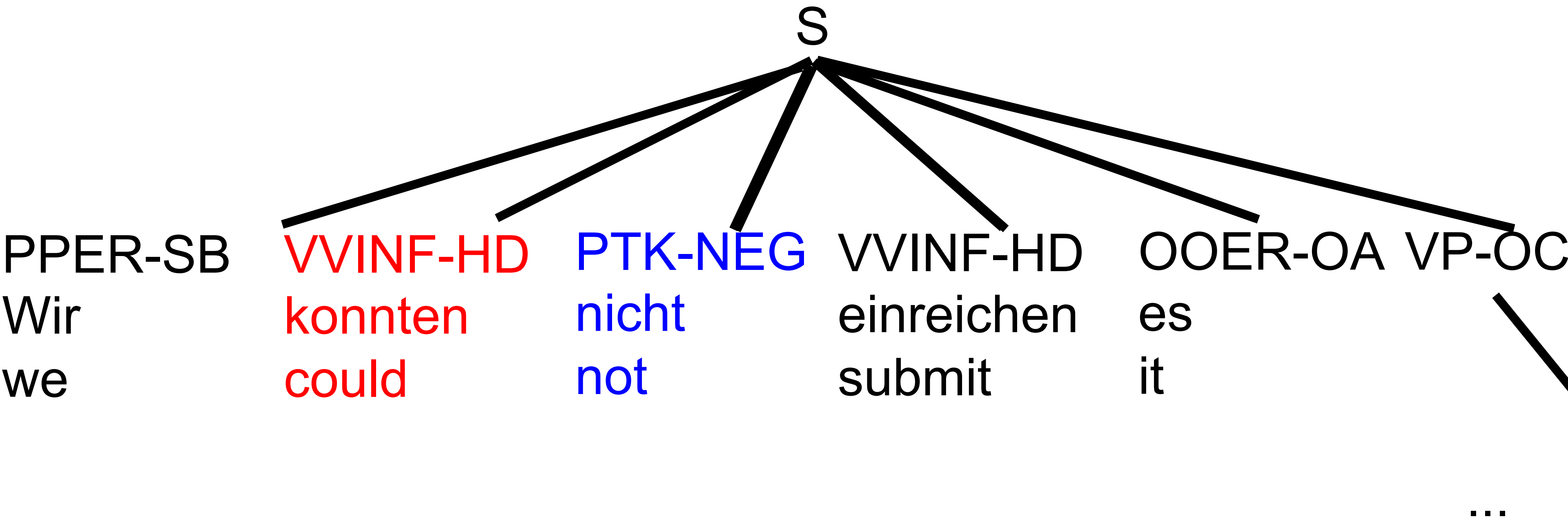
Rule 6: Negation

Negative particle is moved to directly follow the finite verb



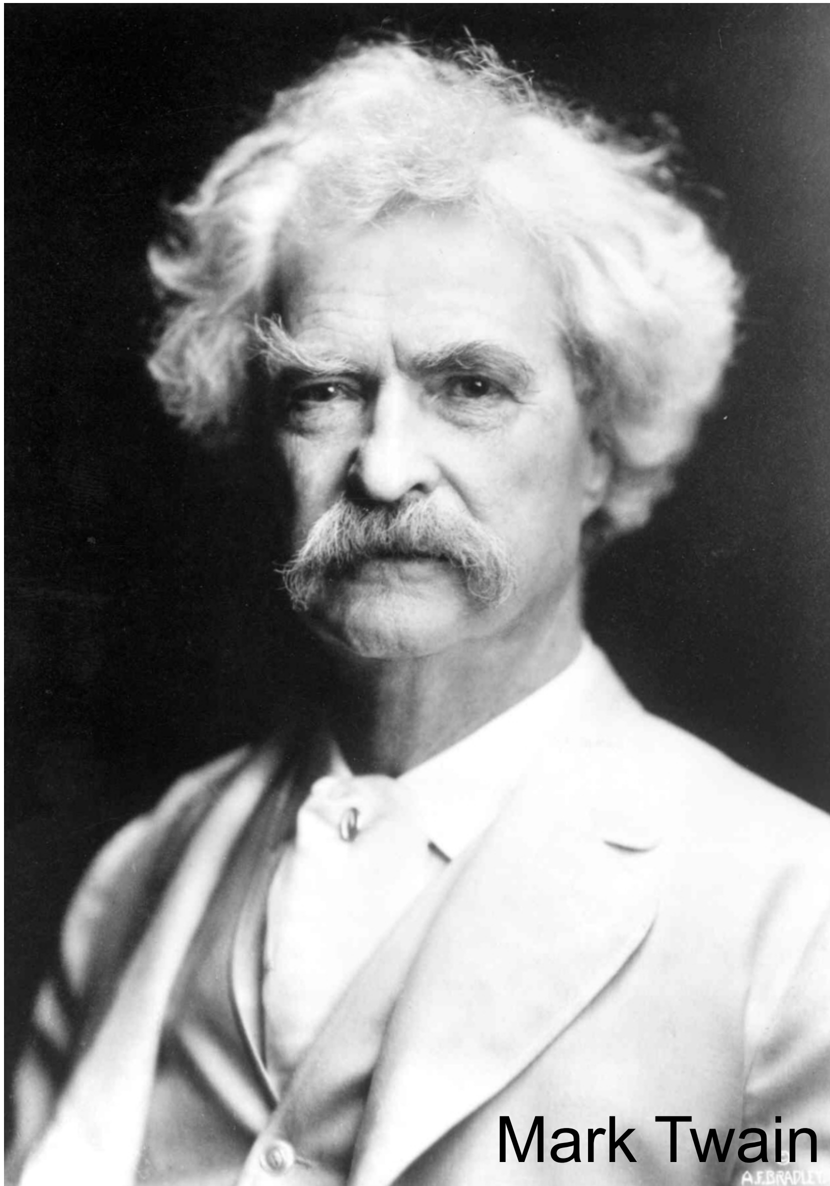
# Clause Restructuring

Rule 6: Negation
Negative particle is moved to directly follow the finite verb



# The Awful German Language

"The Germans have another kind of parenthesis, which they make by splitting a verb in two and putting half of it at the beginning of an exciting chapter and the OTHER HALF at the end of it. Can any one conceive of anything more confusing than that? These things are called 'separable verbs.' The wider the two portions of one of them are spread apart, the better the author of the crime is pleased with his performance."



Mark Twain

# A Less Awful German Language

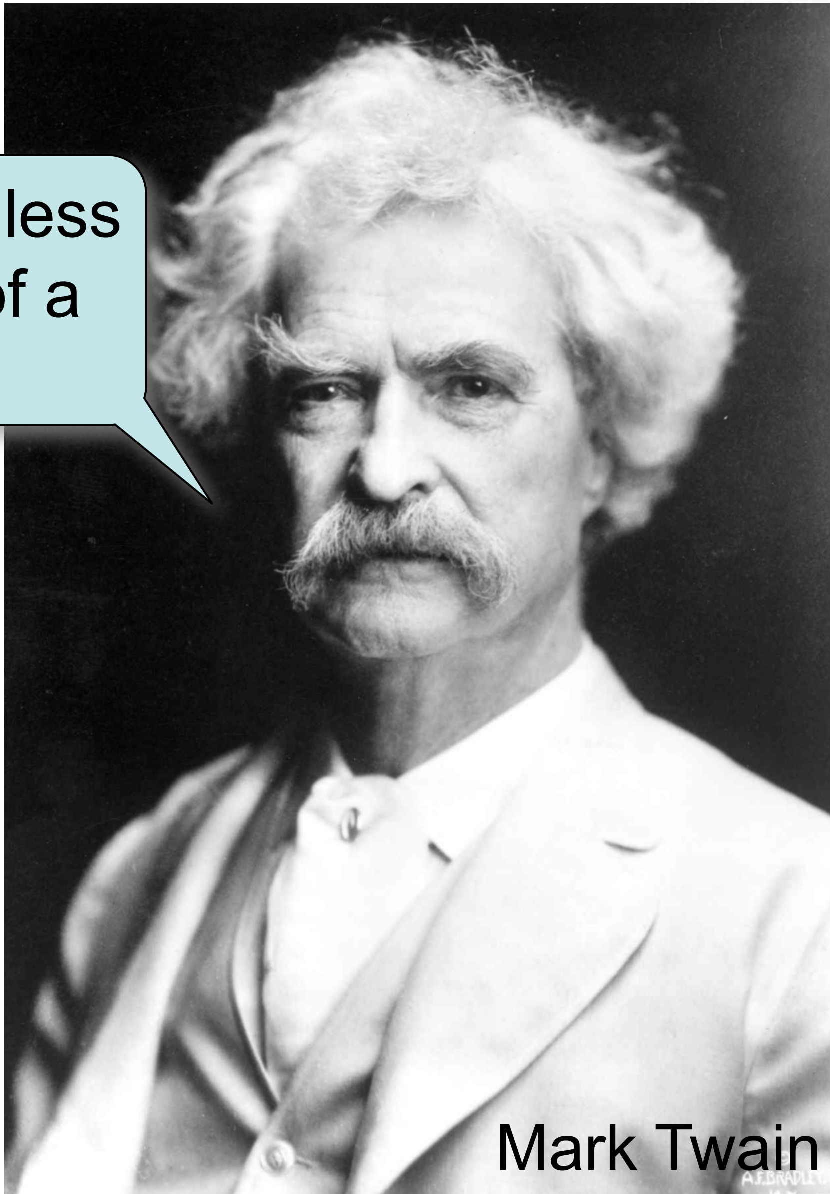
Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen

Now that seems less like the ravings of a madman.

Ich werde aushaendigen Ihnen den Report, damit Sie koennen uebernehmen den eventuell.

I will to_you the report pass_on, so_that you it perhaps adopt can.

I will pass_on to_you the report, so_that you can adopt it perhaps .

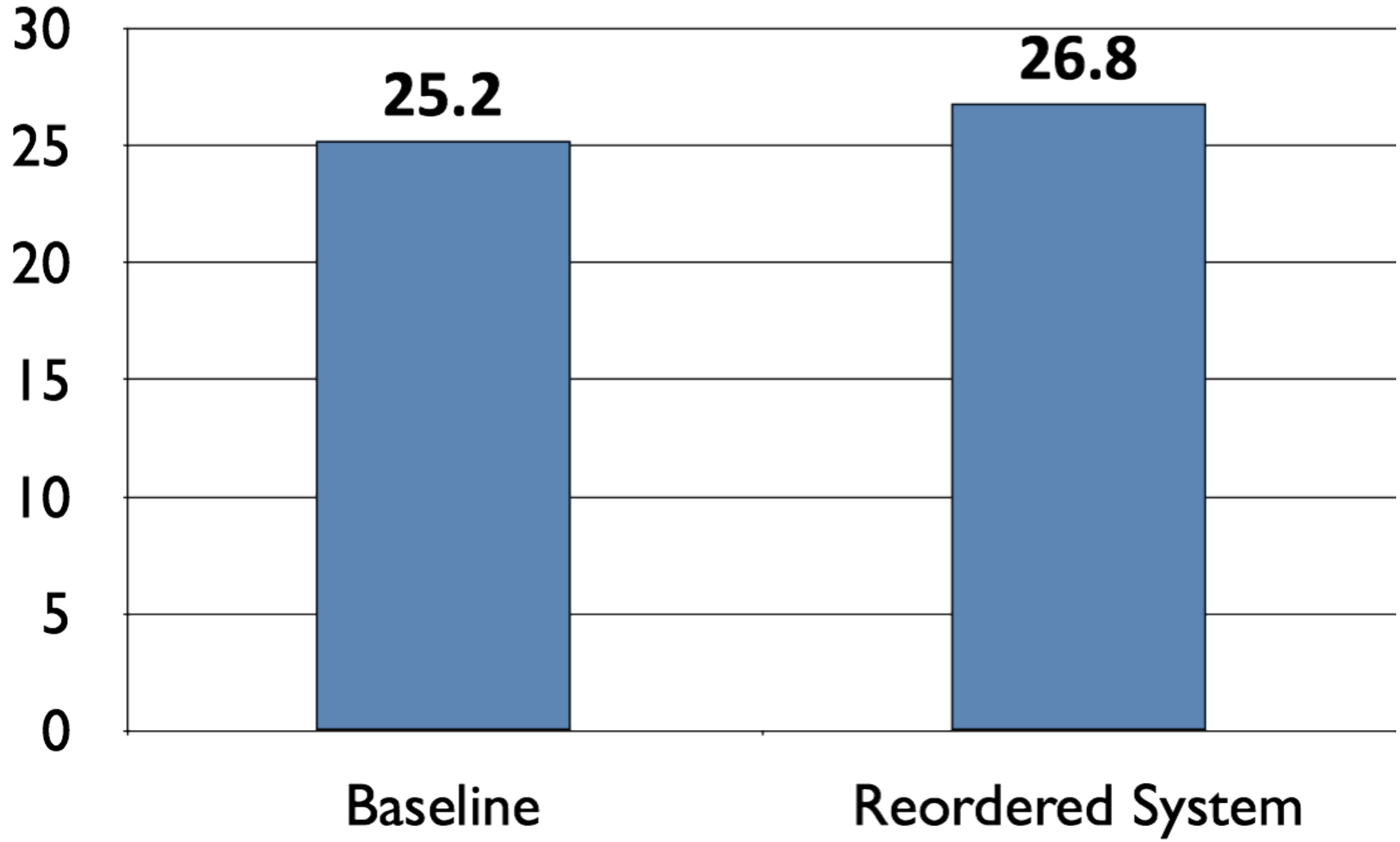


Mark Twain

# Experiments

- Parallel training data: Europarl corpus (751k sentence pairs, 15M German words, 16M English)
  - Parsed German training sentences
  - Reordered the German training sentences with their 6 clause reordering rules
  - Trained a phrase-based model
  - Parsed and reordered the German test sentences
  - Translated them
- Compared against the standard phrase-based model without parsing/reordering

# Bleu score increase




Significant improvement at $p<0.01$ using the sign test

# Human Translation Judgments

- 100 sentences (10-20 words in length)
- Two annotators
- Judged two different versions
  - Baseline system's translation
  - Reordering system's translation
- Judgments: Worse, better or equal
- Sentences were chosen at random, systems' translations were presented in random order

# Human Translation Judgments

| | + | = | – |
|---|---|---|---|
| Annotator 1 | 40% | 40% | 20% |
| Annotator 2 | 44% | 37% | 19% |

+ = reordered translation better
– = baseline better
= = equal

# Examples

| | |
|---|---|
| Reference | I think it is wrong in principle to have such measures in the European Union |

I believe that it is wrong in principle to take such measures in the European Union

I believe that it is wrong in principle, such measure in the European Union to take.

# Examples

| | |
|---|---|
| Reference | The current difficulties should encourage us to redouble our efforts to promote coorperation in the Euro-Mediterranean framework. |
| | The current problems should spur us, our efforts to promote coorperation within the framework of the e-prozesses to be intensified. |
| | The current problems should spur us to intensify our efforts to promote cooperation within the framework of the e-prozesses. |

# Examples

| | |
|---|---|
| Reference | To go on subsidizing tobacco cultivation at the same time is a downright contridiction. |

At the same time, continue to subsidize tobacco growing, it is quite schizophrenic.

At the same time, to continue to subsidize tobacco growing is schizophrenic.

# Examples

| | |
|---|---|
| Reference | We have voted against the report by Mrs. Lalumiere for reasons that include the following: |

We have voted, amongst other things, for the following reasons against the report by Mrs. Lalumiere:

We have, among other things, for the following reasons against the report by Mrs. Lalumiere voted:

# Discussion: Clause Restructuring

- Are you convinced that German-English translation has improved?
- Do you think that this is a good fit for phrase-based machine translation?
- What limitations does this method have?
- (Discuss with your neighbor.)

# Limitations

- Requires a parser for the source language
  - We have parsers for only a small number of languages
  - Penalizes “low resource languages”
  - Fine for translating from English into other languages
- Involves hand crafted rules
- Removes the nice language-independent qualities of statistical machine translation