Syntax-based Translation

Part 1: Re-ordering for Phrase-based translation

Machine Translation
Lecture 12
**Instructor: Chris Callison-Burch**

Thanks to Michael Collins for many of today's slides.
Take a look at Mike's course: http://www.cs.columbia.edu/~cs4705/
and his Coursera course: https://class.coursera.org/nlangp-001

Goals

- Understand why syntax is important for reordering models

- Review non-syntactic reordering models for phrase-based machine translation

- Review the "Clause Restructuring" approach of Collins, Koehn, and Kucerova

- Understand why it is a good fit for phrase-based machine translation

- Discuss its limitations

Phrase-based model
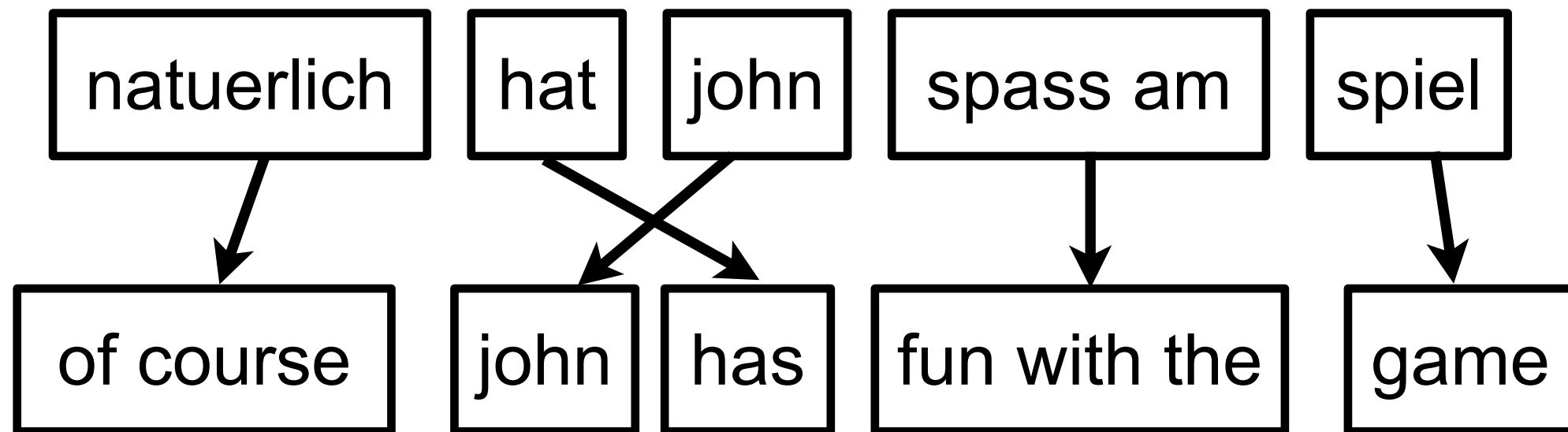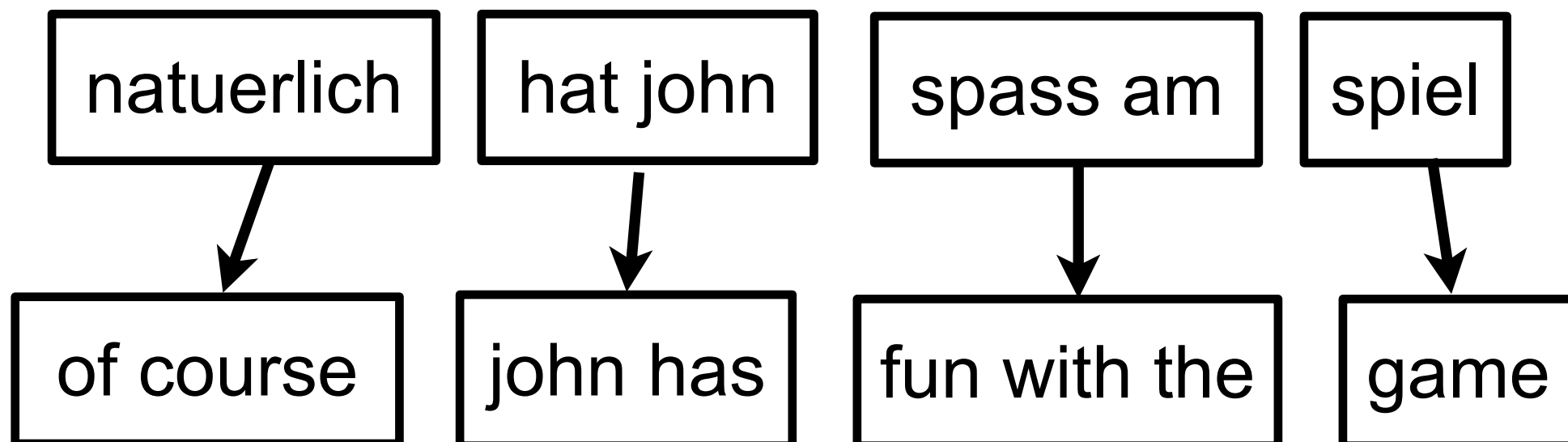

- Foreign input is segmented in phrases
- Each phrase is translated into English
- Phrases are reordered

Some Reordering Already Captured
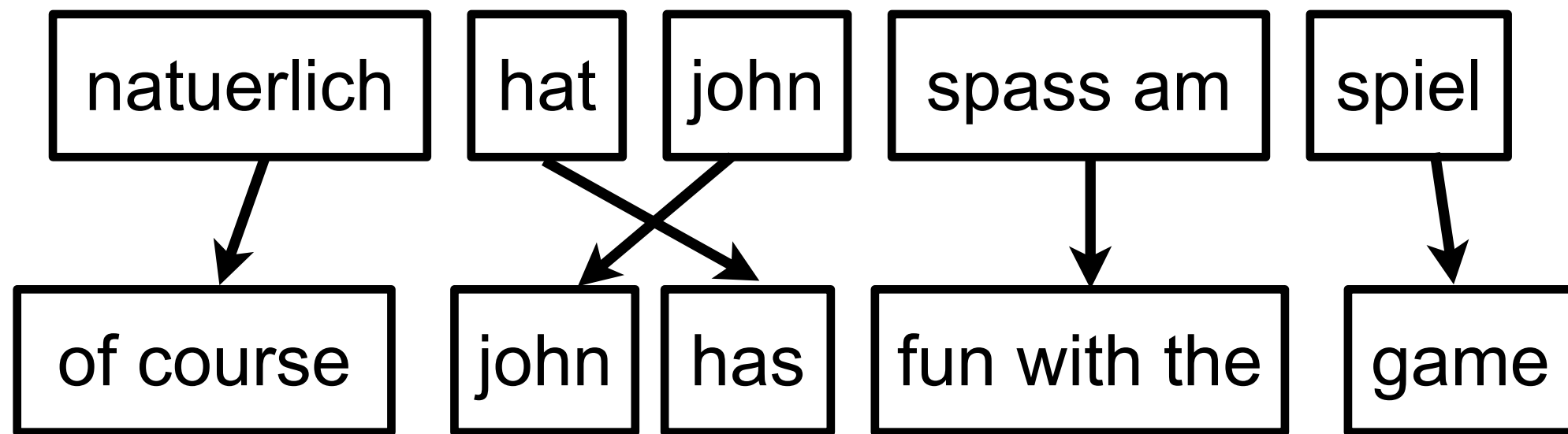
- Local reordering can be captured within phrases

Phrase translation table

- Main knowledge source: table with phrase translations and their probabilities

- Example: phrase translations for natuerlich

Source	Translation	Probability $\phi(e|f)$
natuerlich	of course	0.5
natuerlich	naturally	0.3
natuerlich	of course ,	0.15
natuerlich	, of course ,	0.05

Probabilistic Model

- Bayes rule

  - $\mathbf{e}_{best} = \arg\max_{\mathbf{e}} \ p(\mathbf{e}|\mathbf{f})$

    $= \arg\max_{\mathbf{e}} \ p(\mathbf{f}|\mathbf{e}) \ p_{lm}(\mathbf{e})$

  - translation model $p(\mathbf{e}|\mathbf{f})$

  - language model $p_{lm}(\mathbf{e})$
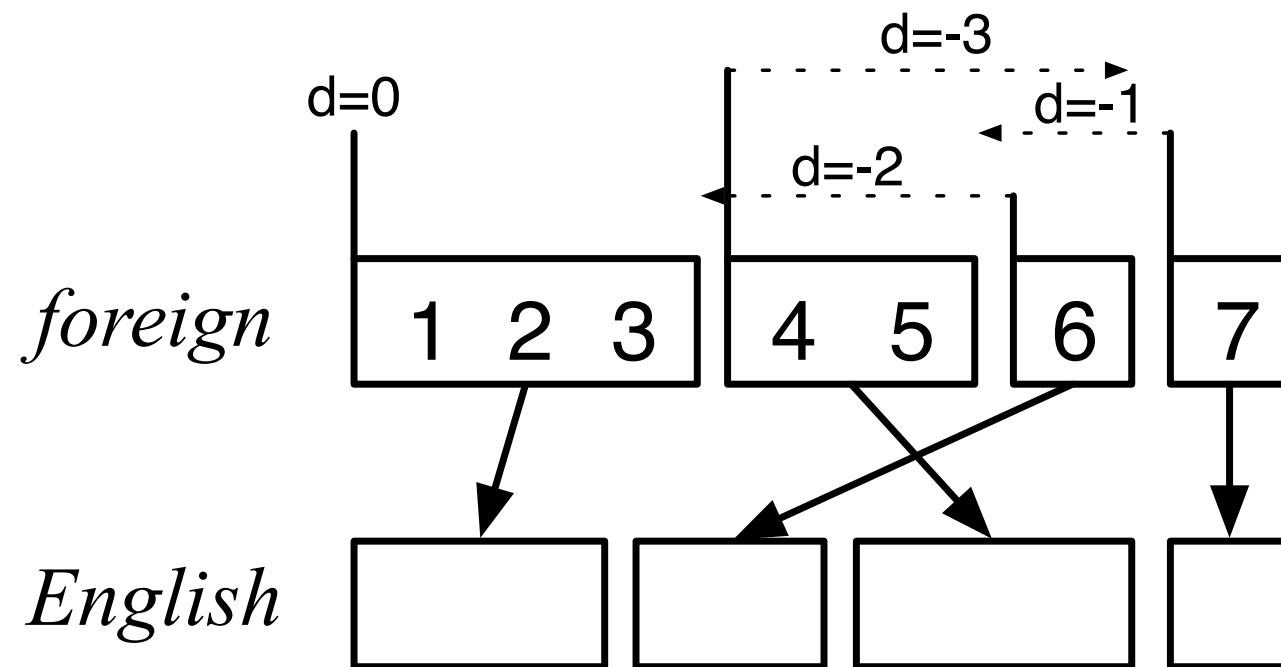
- Reordering score can be incorporated in the TM

$$p(\bar{f}_1^I|\bar{e}_1^I) = \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i) \ d(start_i - end_{i-1} - 1)$$

  - phrase translation probability ф

  - reordering probability d

Log-linear model

$$p(e, a|f) = \exp(\lambda_\phi \sum_{i=1}^{I} \log \phi(\bar{f}_i|\bar{e}_i)+$$

$$\lambda_d \sum_{i=1}^{I} \log d(a_i - b_{i-1} - 1)+$$

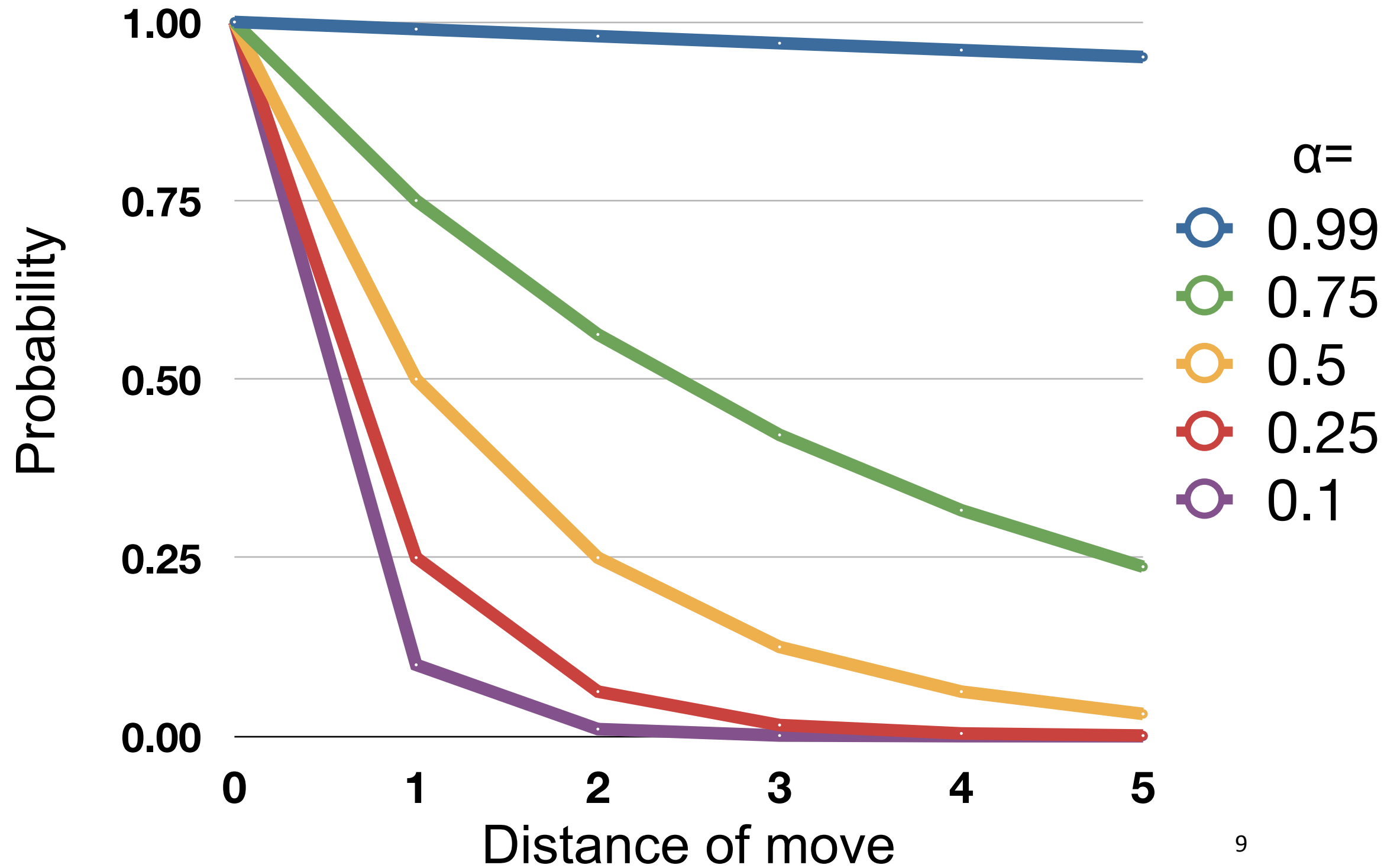$$\lambda_{LM} \sum_{i=1}^{|\mathbf{e}|} \log p_{LM}(e_i|e_1...e_{i-1}))$$

Distance-based Reordering

phrase	translates	movement	distance
1	1–3	start at beginning	0
2	6	skip over 4–5	+2
3	4–5	move back over 4–6	-3
4	7	skip over 6	+1

Scoring function: $d(x) = \alpha^{|x|}$ – exponential with distance

Values of α

Discussion: Distance-based reordering

- What do you think of it?

- Is it a good model for how reordering works across languages?

- What is it missing?

(Discuss with your neighbor)

Distance-based reordering

- Small values of α, severely discourage reordering
  - Limit reordering to monotonic or a narrow window
  - OK for languages with very similar word orders
  - Bad for languages with different word orders
- The distance-based penalty applies uniformly to all words and all word types
  - Doesn't know that adjectives and nouns should swap when translating from French to English
- Puts most responsibility on the language model

How else could we model reordering?

- Why not assign a distinct reordering probability to each word/phrase in the phrase table?
  - p(reorder | f, e)

- This is known as lexicalized reordering

- How can we estimate that probability?

Lexicalized Reordering model

m: monotone (keep order)
s: swap order
d: become discontinuous

Reordering features are
probability estimates of s,
d, and m

Lexicalized Reordering table

- Identical phrase pairs <f,e> as in the phrase translation table

- Contains values for p(monotone|e,f), p(swap|e,f), p(discontinuous|e,f)

| Source	Translation	p(m|e,f)	p(s|e,f)	p(d|e,f)
natuerlich	of course	0.52	0.08	0.4
natuerlich	naturally	0.42	0.1	0.48
natuerlich	of course ,	0.5	0.001	0.499
natuerlich	, of course	0.27	0.17	0.56

Discussion: Is this better?

- Do you think that this is a more sensible reordering model than the distance-based one?

- How could you determine if it is better or not?

- What do you think that it still lacks?
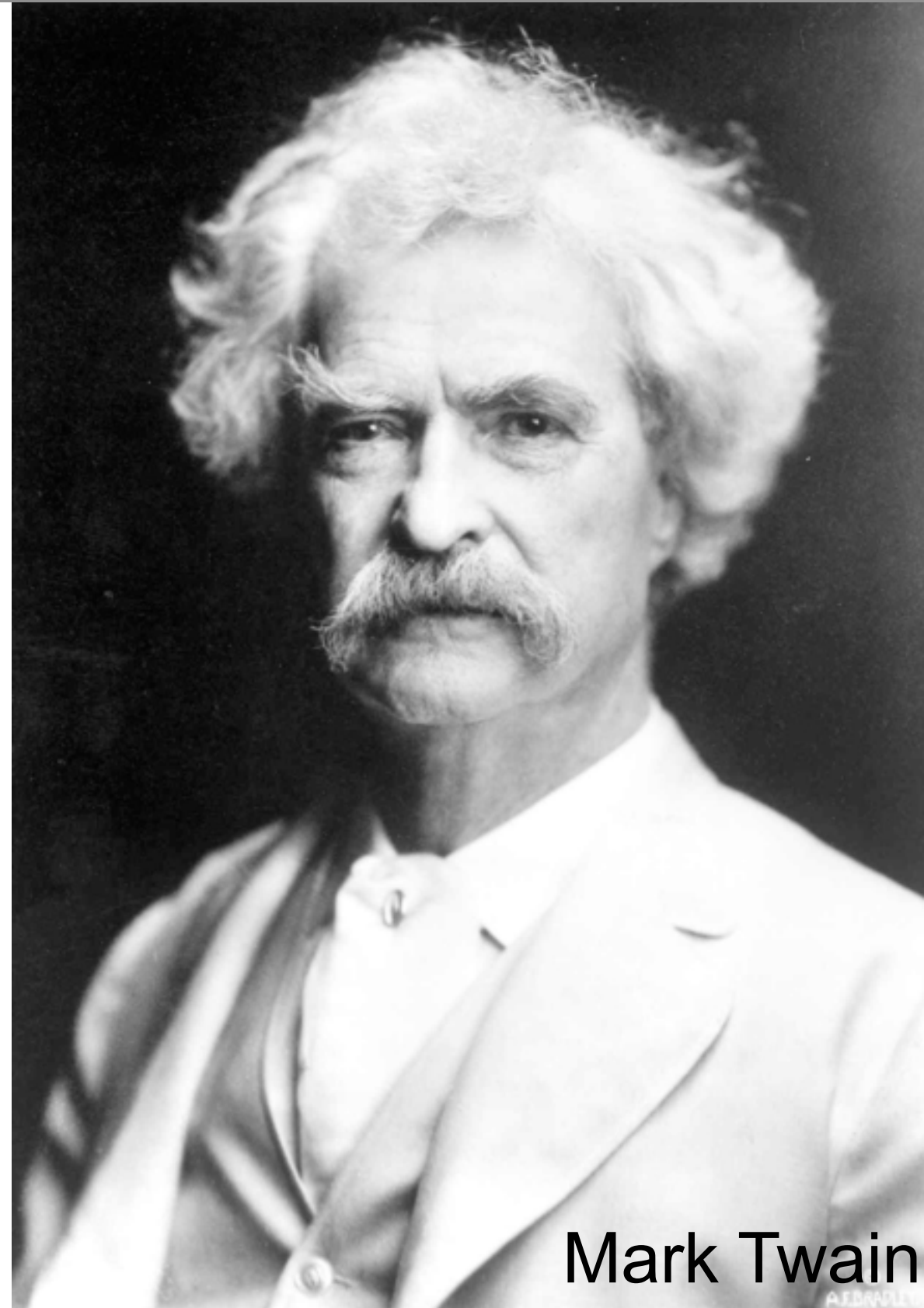
  (Discuss with your neighbor)

Empirically, yes!

Koehn et al, IWSLT 2005
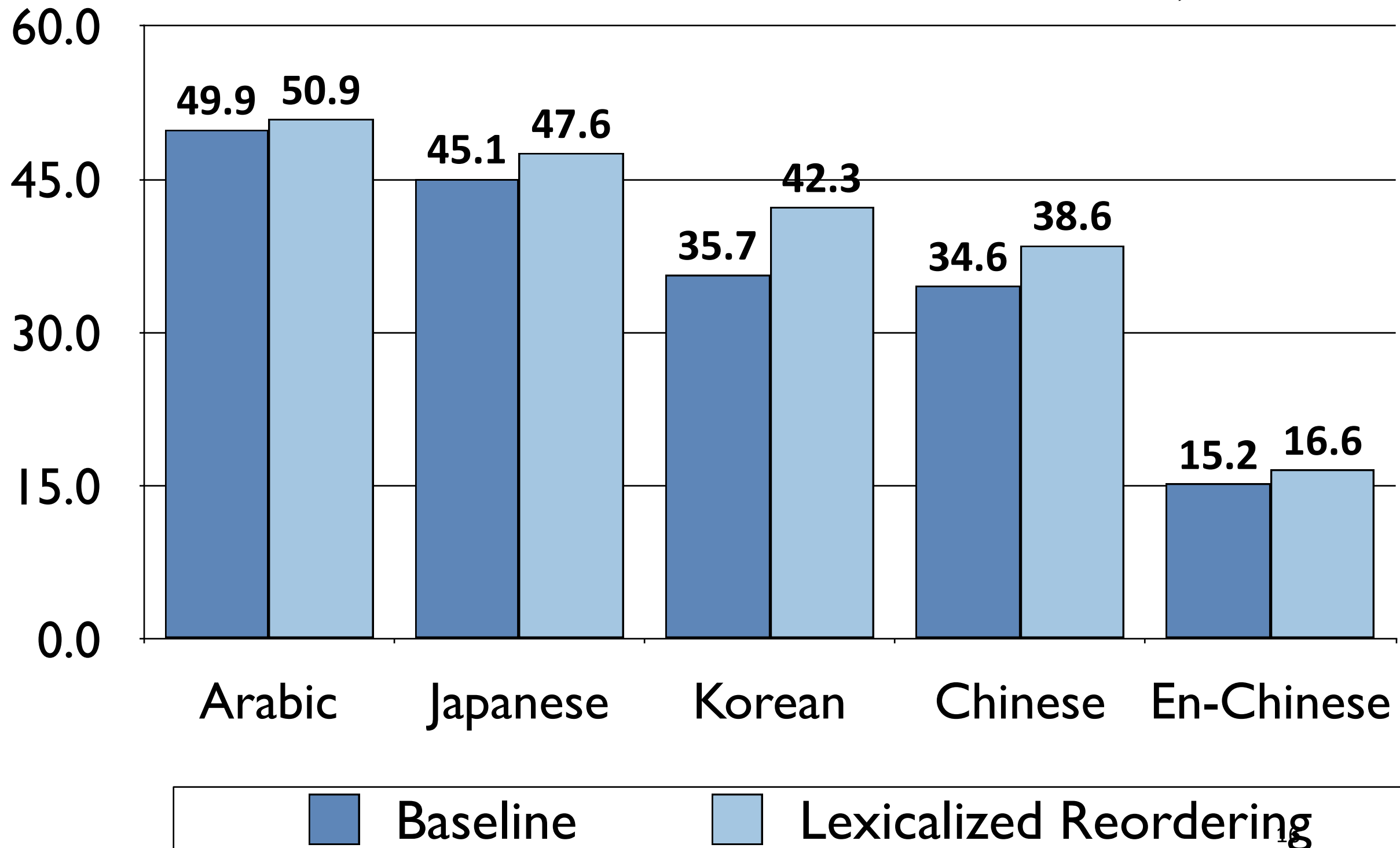
The Awful German Language

"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

German verbs

Ich werde Ihnen den Report aushaendigen .
I   will   to_you the   report   pass_on   .

Ich werde Ihnen   die entsprechenden Anmerkungen aushaendigen .
I   will   to_you the   corresponding comments   pass_on   .

Ich werde Ihnen die entsprechenden Anmerkungen am Dienstag aushaendigen
I   will   to_you the   corresponding comments   on Tuesday   pass_on

German free word order

The finite verb always appears in 2nd position, but
Any constituent (not just the subject) can appear in the 1st position

I   will   to_you the report   pass_on

To_you   will   I the report   pass_on

The report will   I to_you pass_on

German verbs

Main clause

Ich	werde	Ihnen	den	Report	aushaendigen	,
I	will	to_you	the	report	pass_on	,

Subordinate clause

damit	Sie	den	eventuell	uebernehmen	koennen	.
so_that	you	it	perhaps	adopt	can	.

Collins' Motivation

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.

Collins' Pre-ordering Model

Step 1: Reorder the source language

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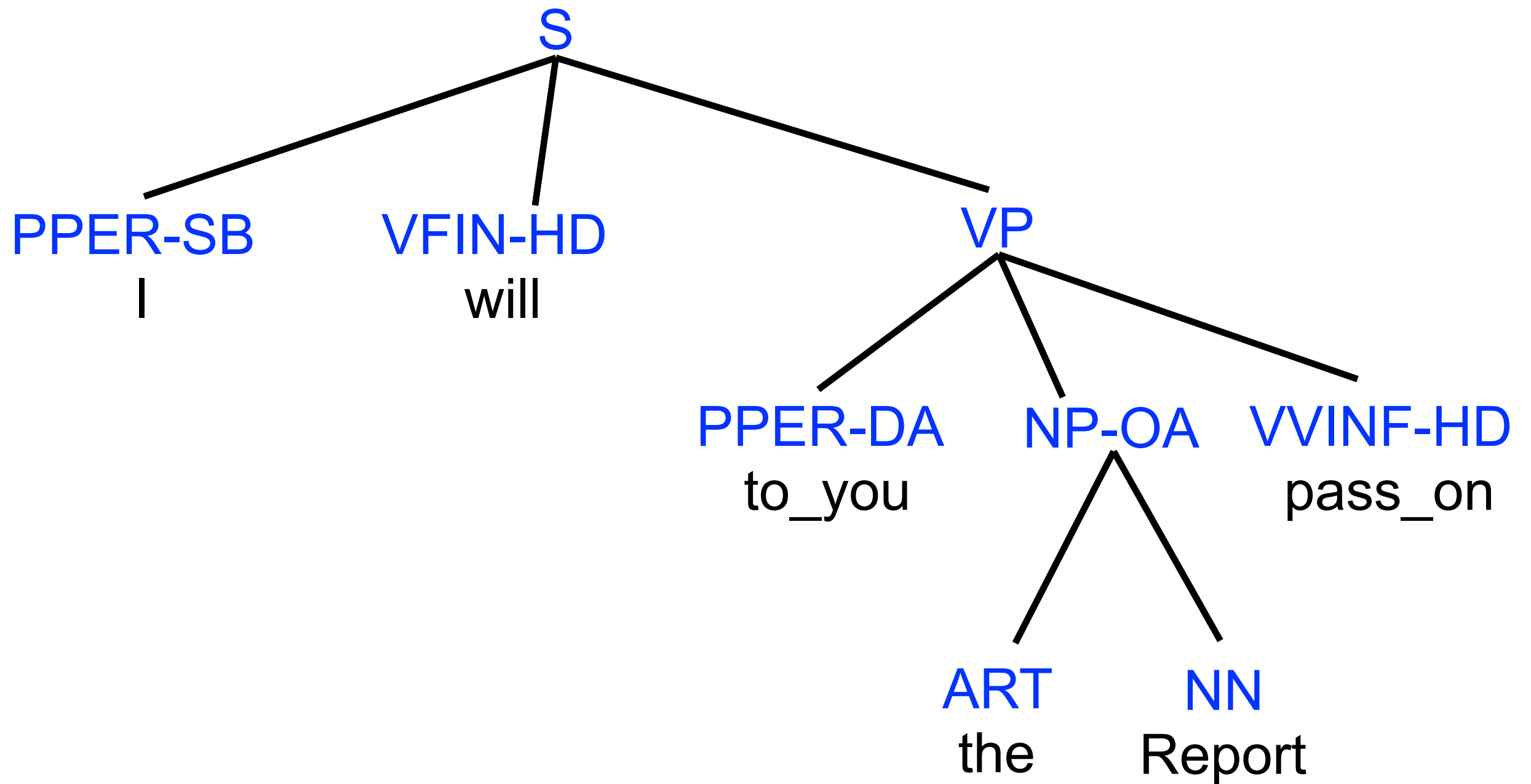damit Sie den eventuell uebernehmen koennen .
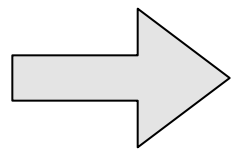
Ich werde aushaendigen Ihnen den Report ,
damit Sie koennen uebernehmen den eventuell .

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

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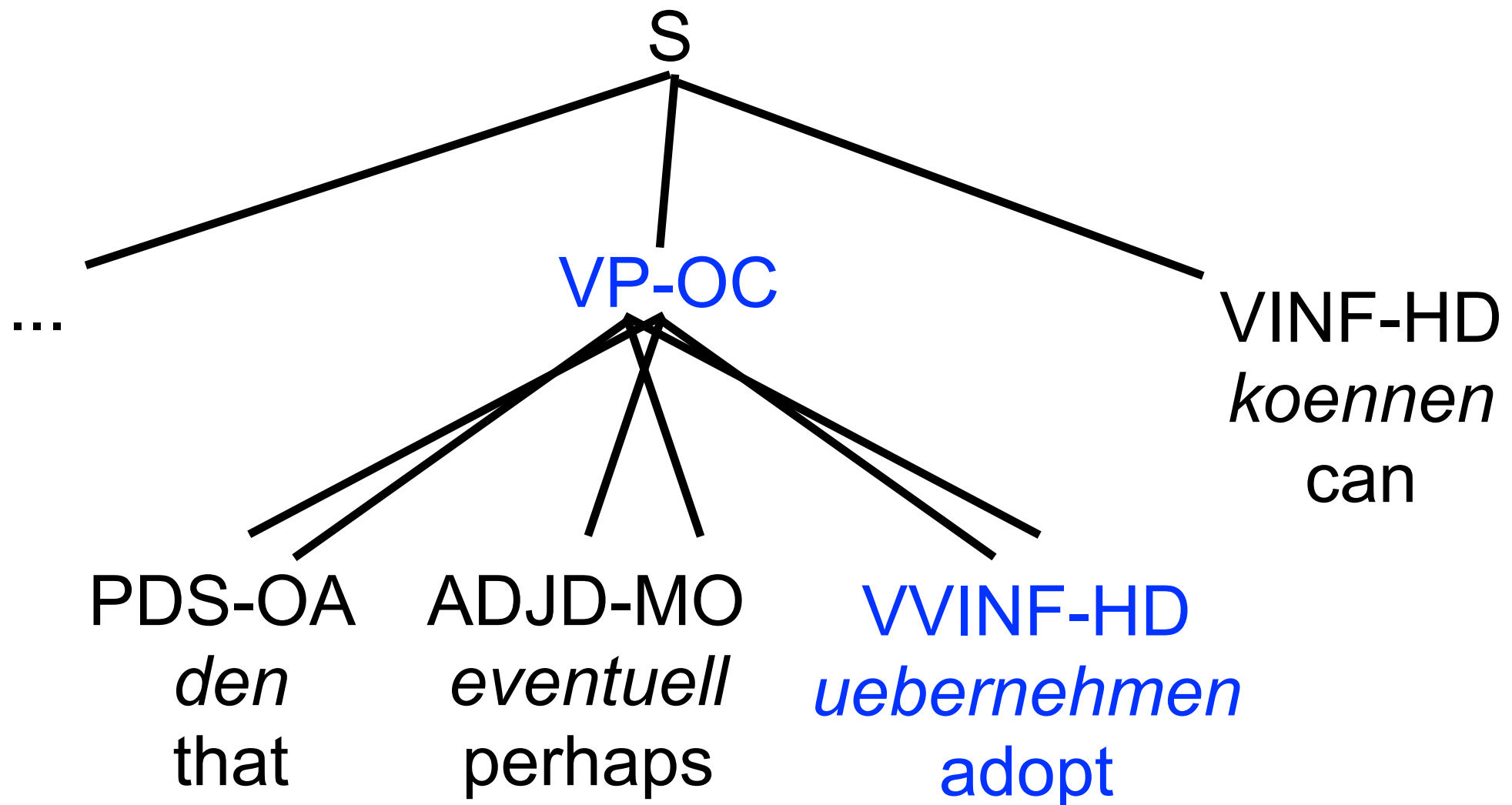to the reordered input.

Example Parse Tree

Clause Restructuring

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

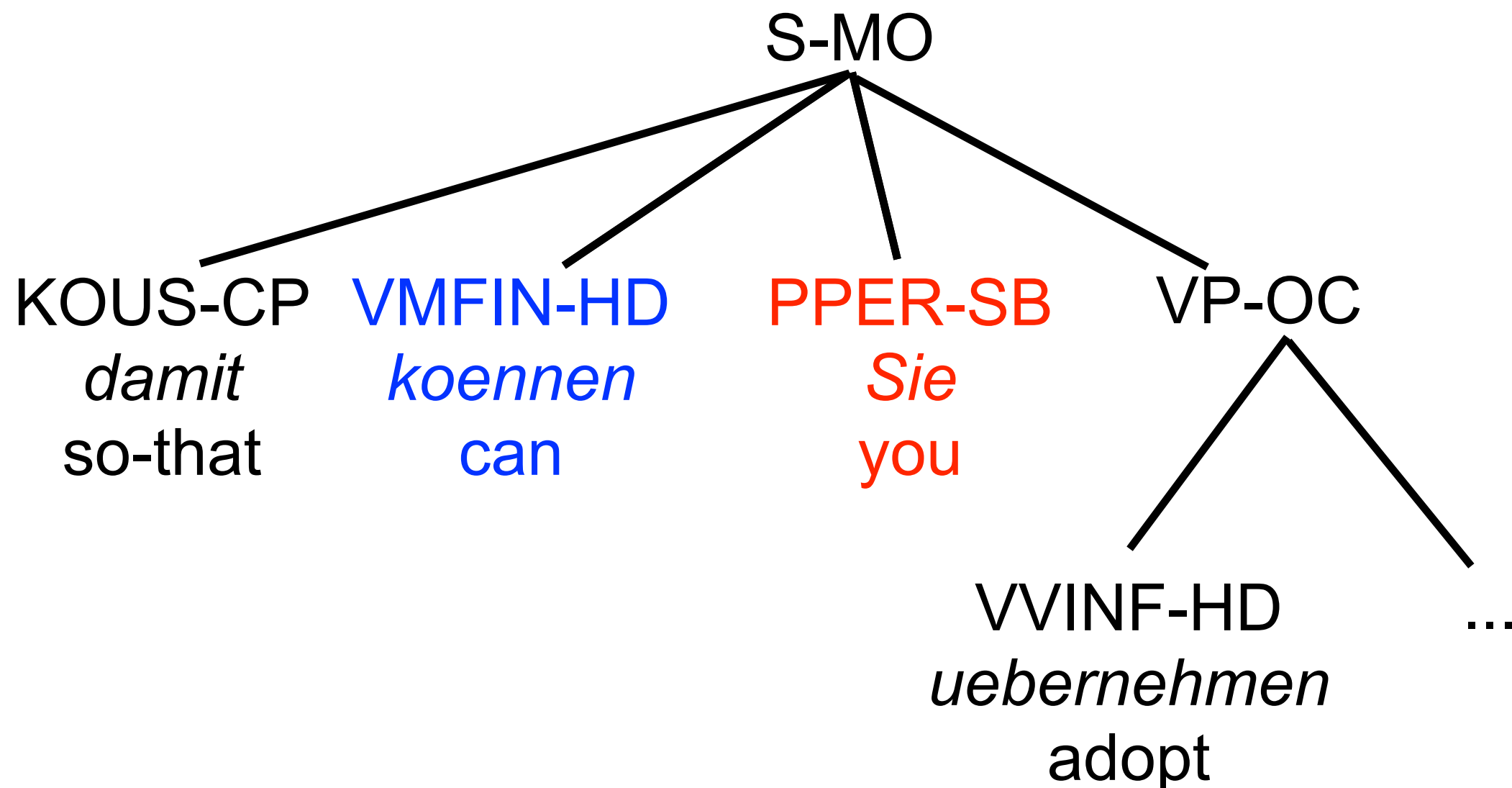# Clause Restructuring

Rule 2: **Verbs follow complementizers**
In a subordinated clause mote the head of the clause to follow the complementizer
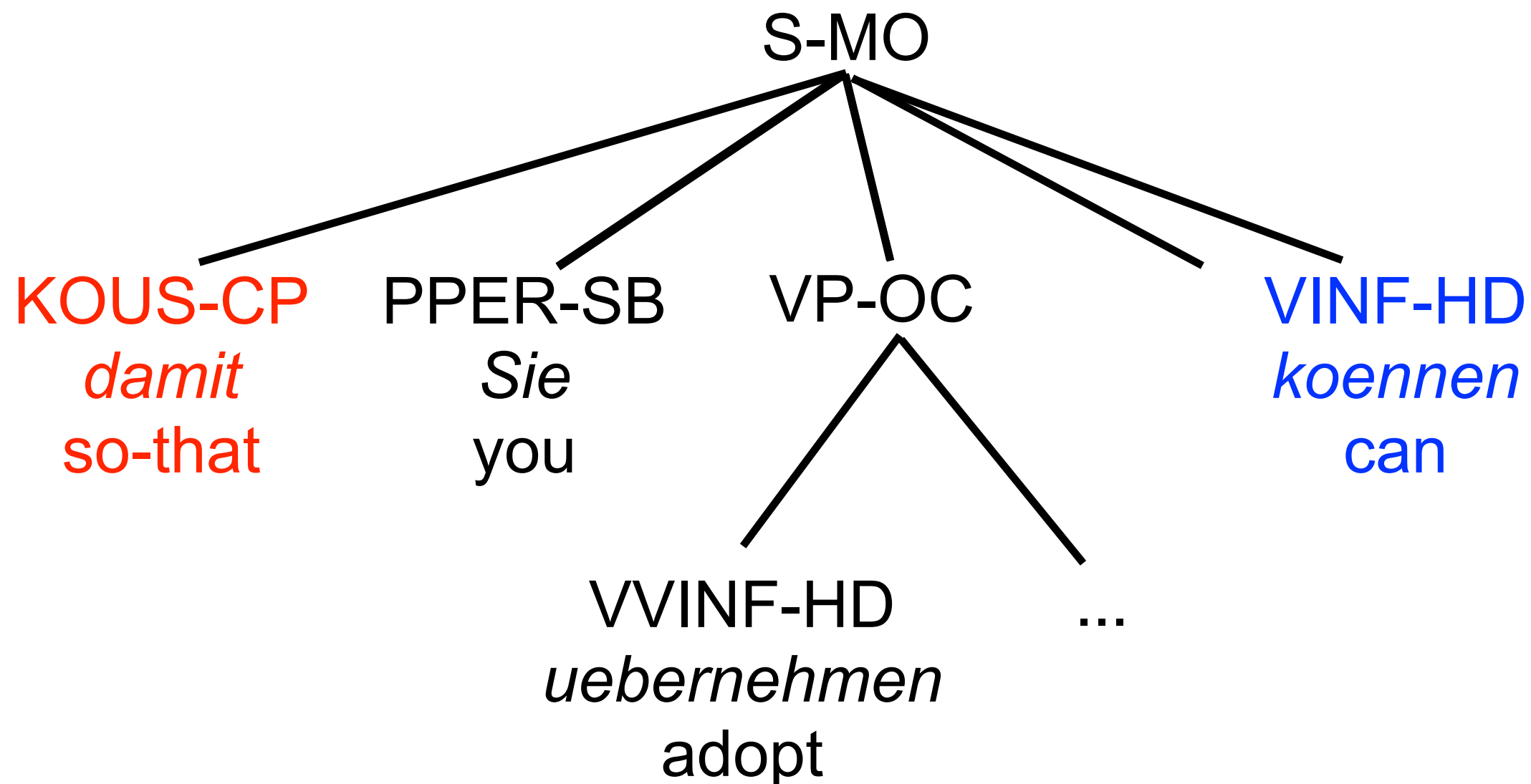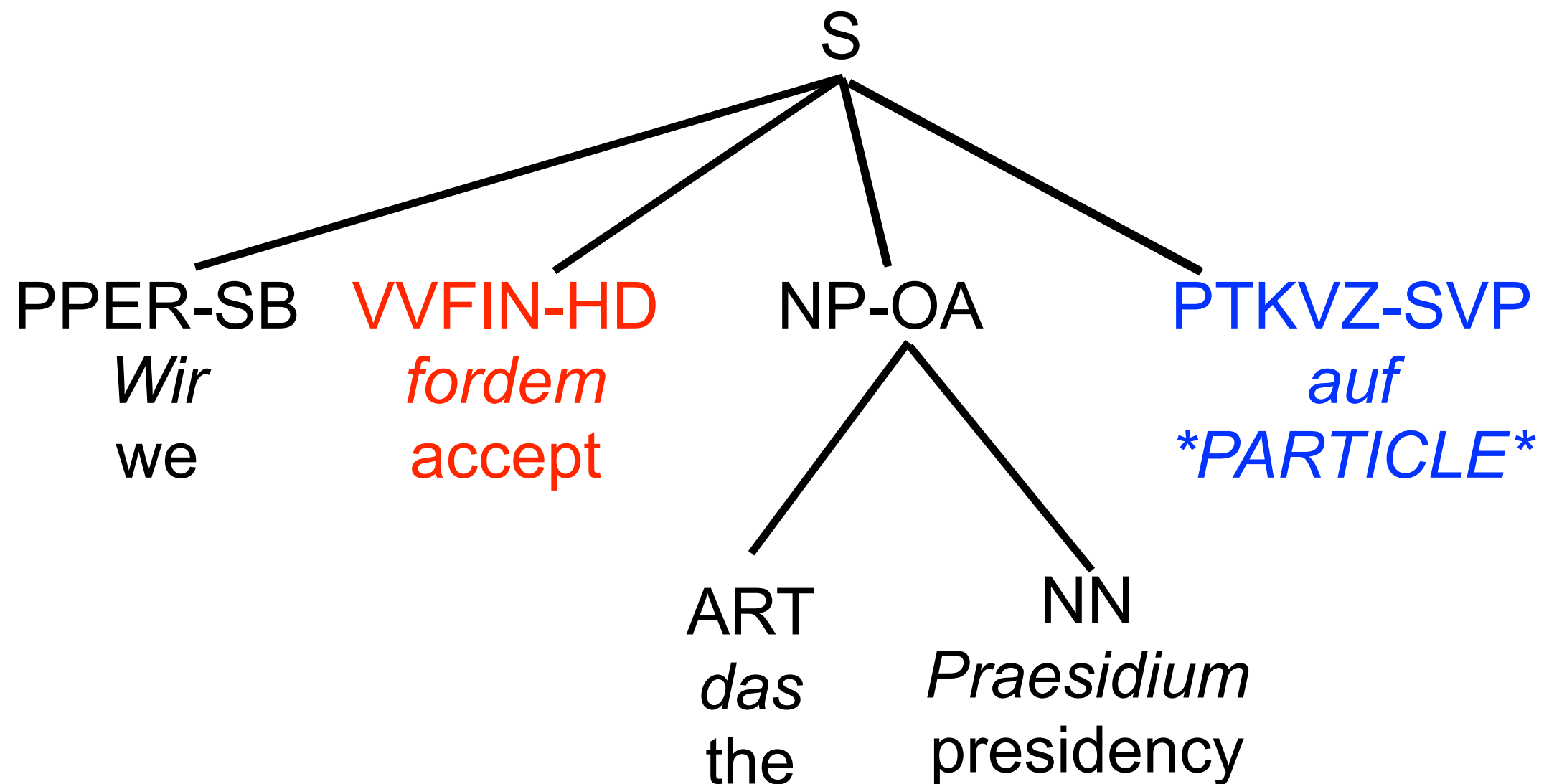
Clause Restructuring

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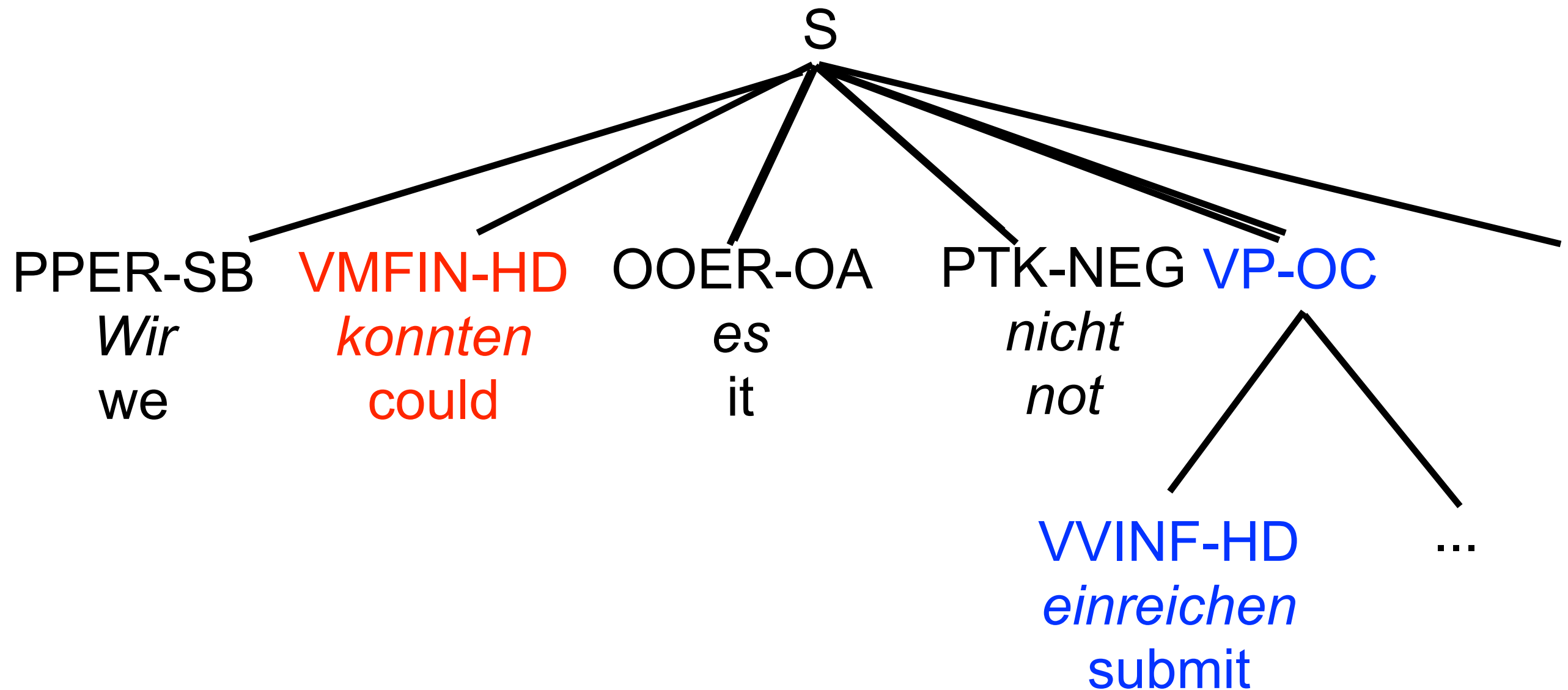
Clause Restructuring

Rule 4: **Particles**

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

Clause Restructuring
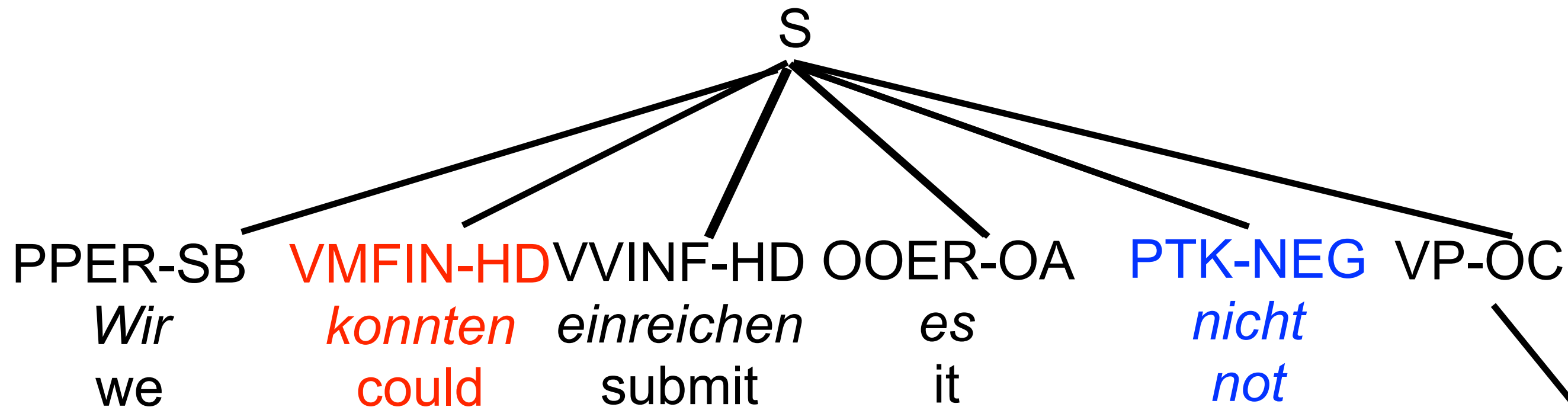
Rule 5: **Infinitives**

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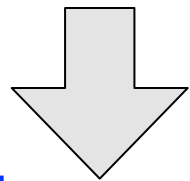
Clause Restructuring

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

Experiments

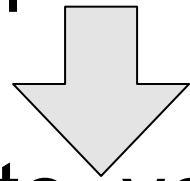- Parallel training data: Europarl corpus (751k sentence pairs, 15M German words, 16M English)
- Parsed German training sentences
- Reordered the German training sentences with their 6 clause reordering rules
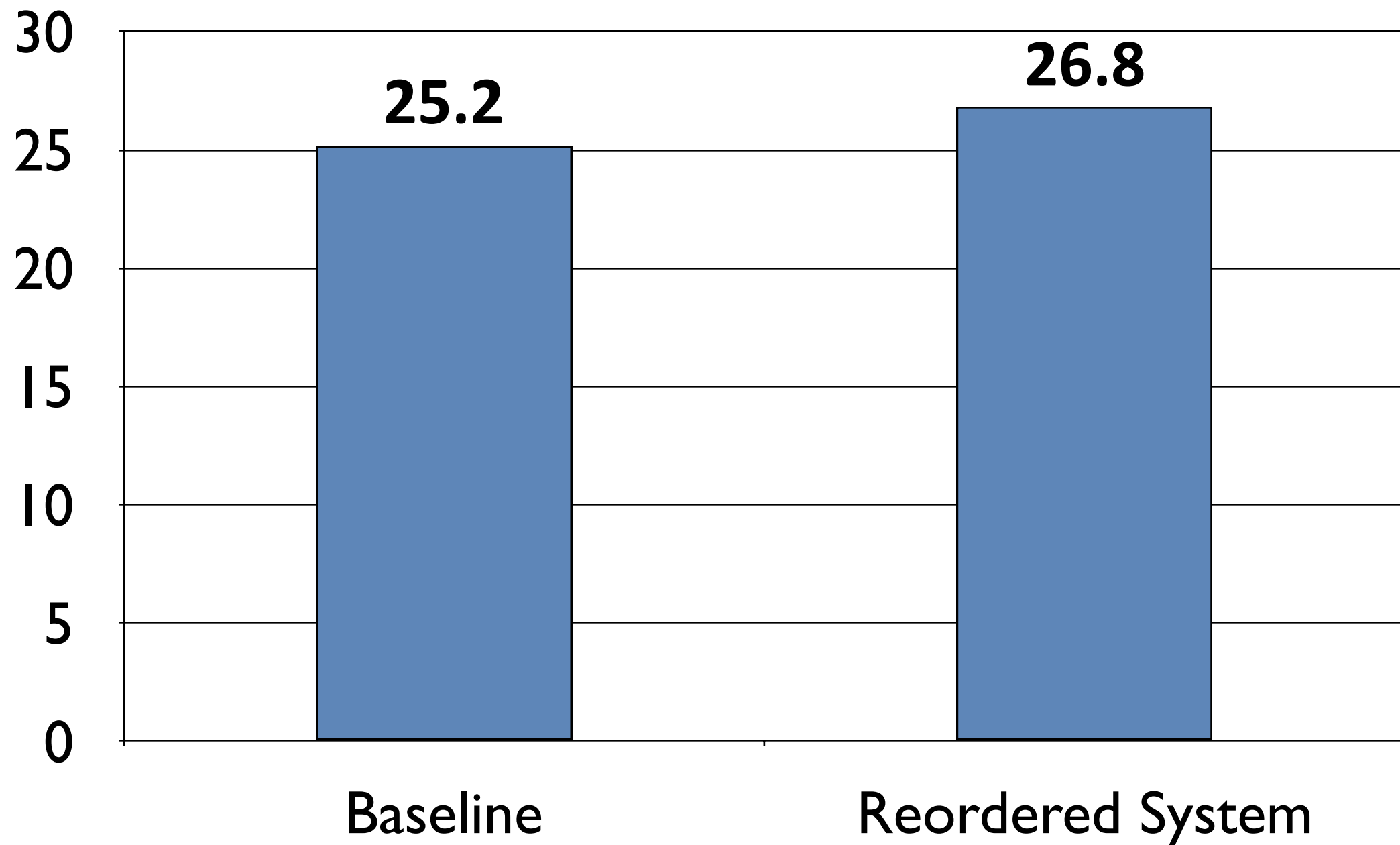- Trained a phrase-based model
- Parsed and reordered the German test sentences
- Translated them
- Compared against the standard phrase-based model without parsing/reordering

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.

| **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

Learning the Rules Automatically

- Great term project idea!
- "Improving a statistical MT system with automatically learned rewrite patterns"

  by Fei Xia and Michael McCord (Coling 2004)

Syntactic LMs

- Our goal is reorder the translated phrases so that they are grammatical English

- Isn't the language model probability supposed to do that already?

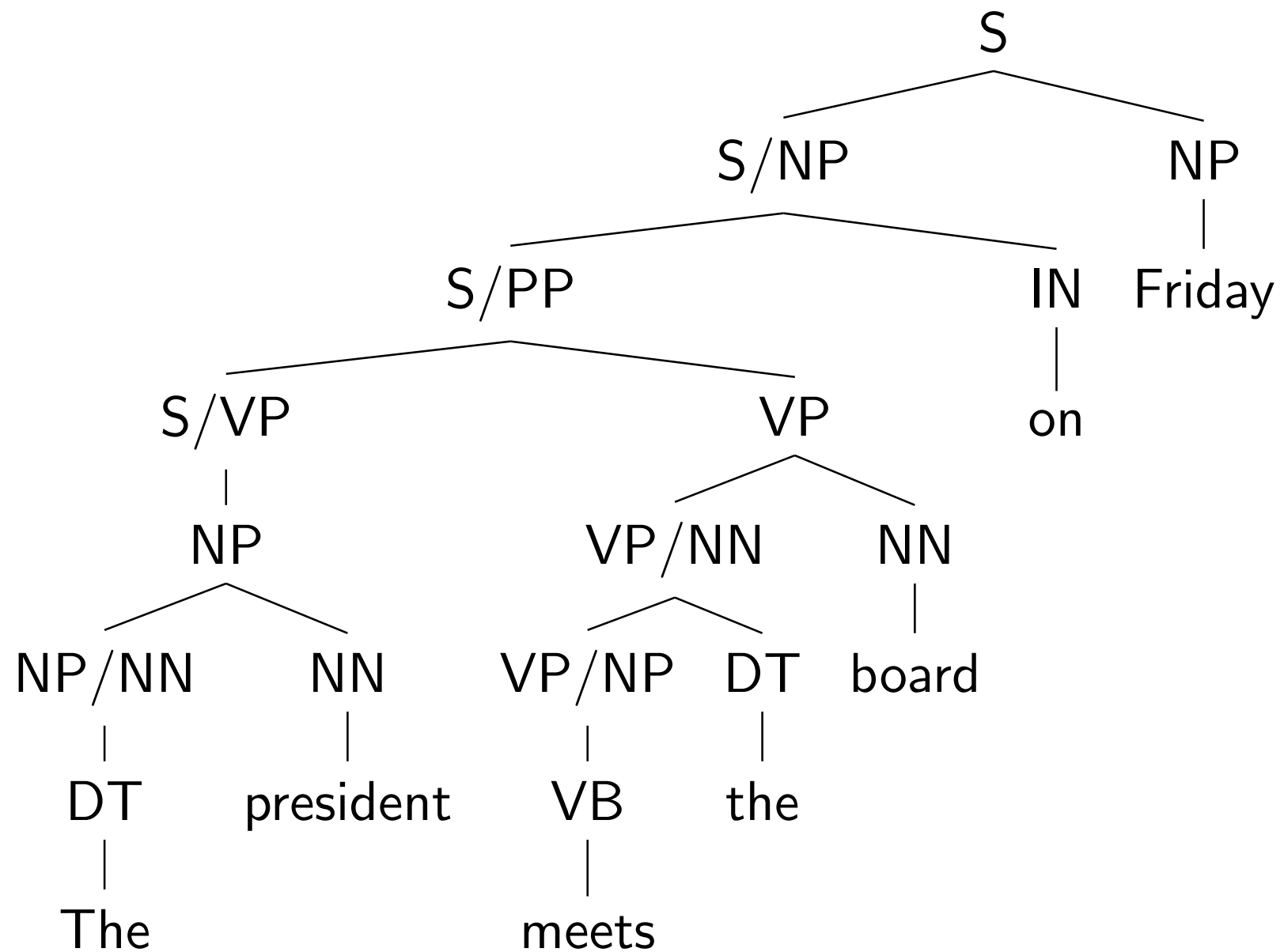- Instead of an n-gram model, could we augment the LM with syntactic information?

Statistical parsing
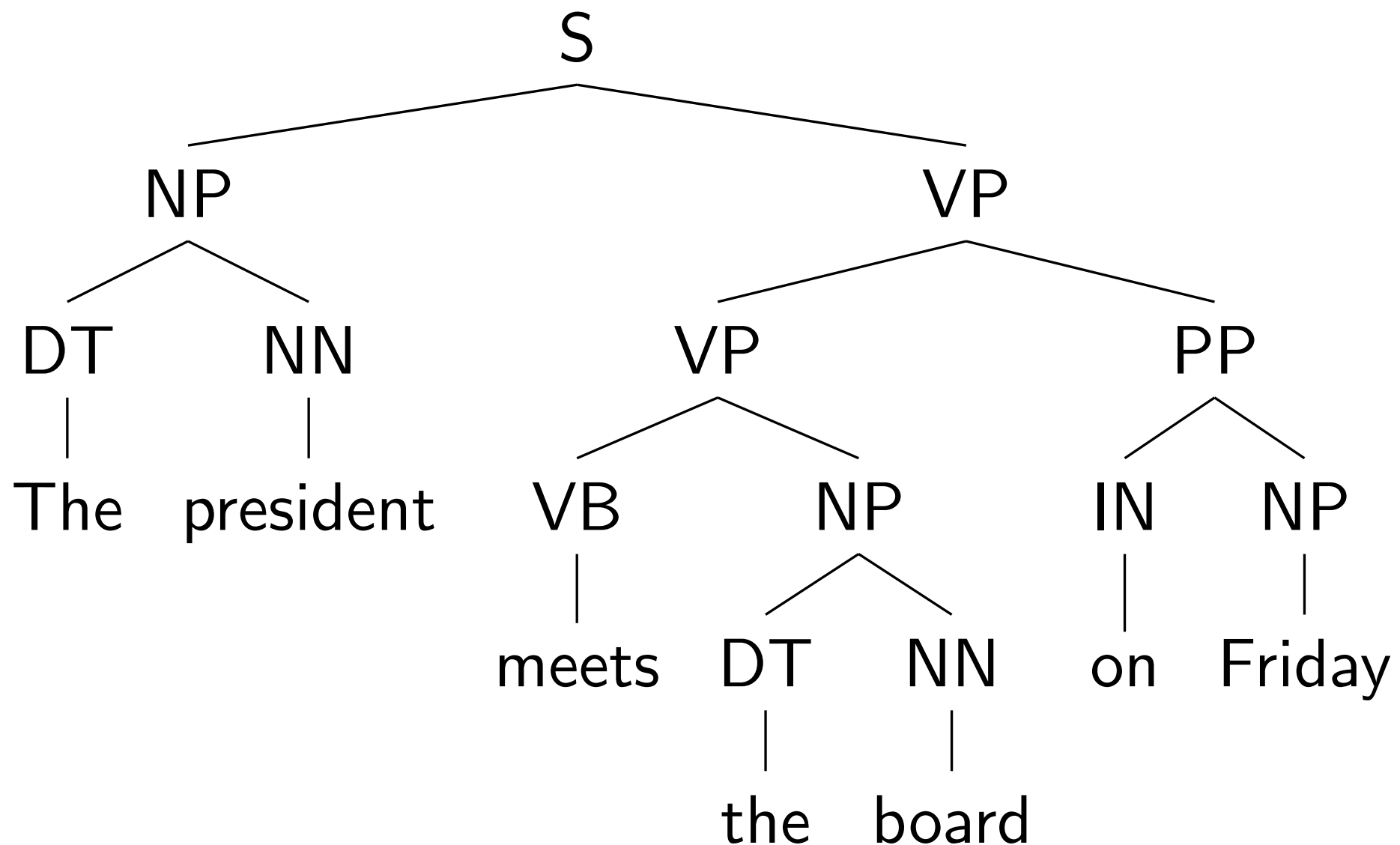
Problem: bottom up parsing requires whole sentence

We need the LM to be able to score partial translations

One possibility: Incremental parsing

More later

- Next, we move away from phrase-based MT and talk synchronous grammar models
- Questions about this material?

Announcements

- HW3 due in a week
- Project guidelines have been posted

The Awful German Language

"Some German words are so long that they have a perspective. Freundschaftsbezeigungen. Dilettantenaufdringlichkeiten. Stadtverordnetenversammlungen. These things are not words, they are alphabetical processions. And they are not rare; one can open a German newspaper at any time and see them marching majestically across the page–and if he has any imagination he can see the banners and hear the music, too."
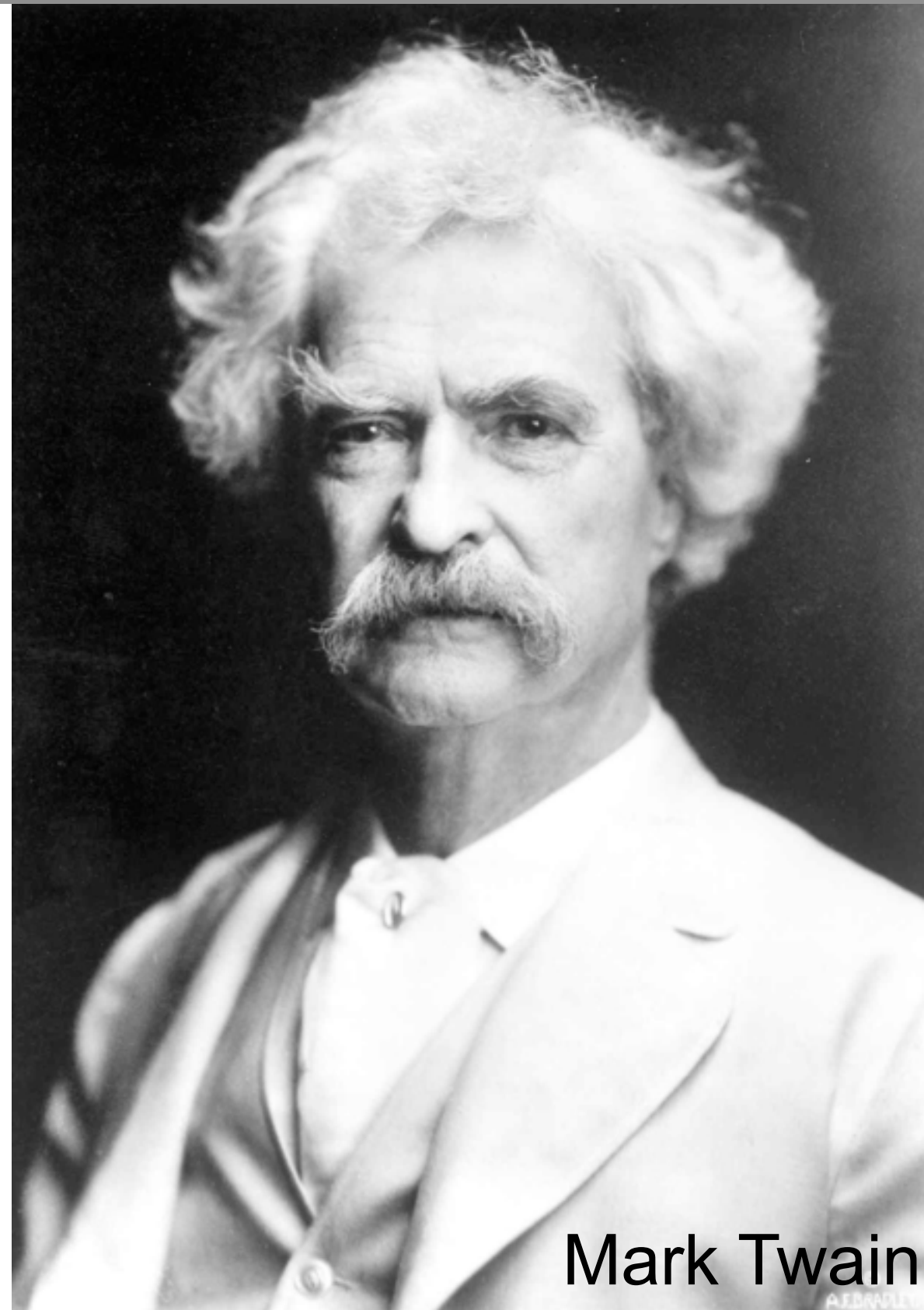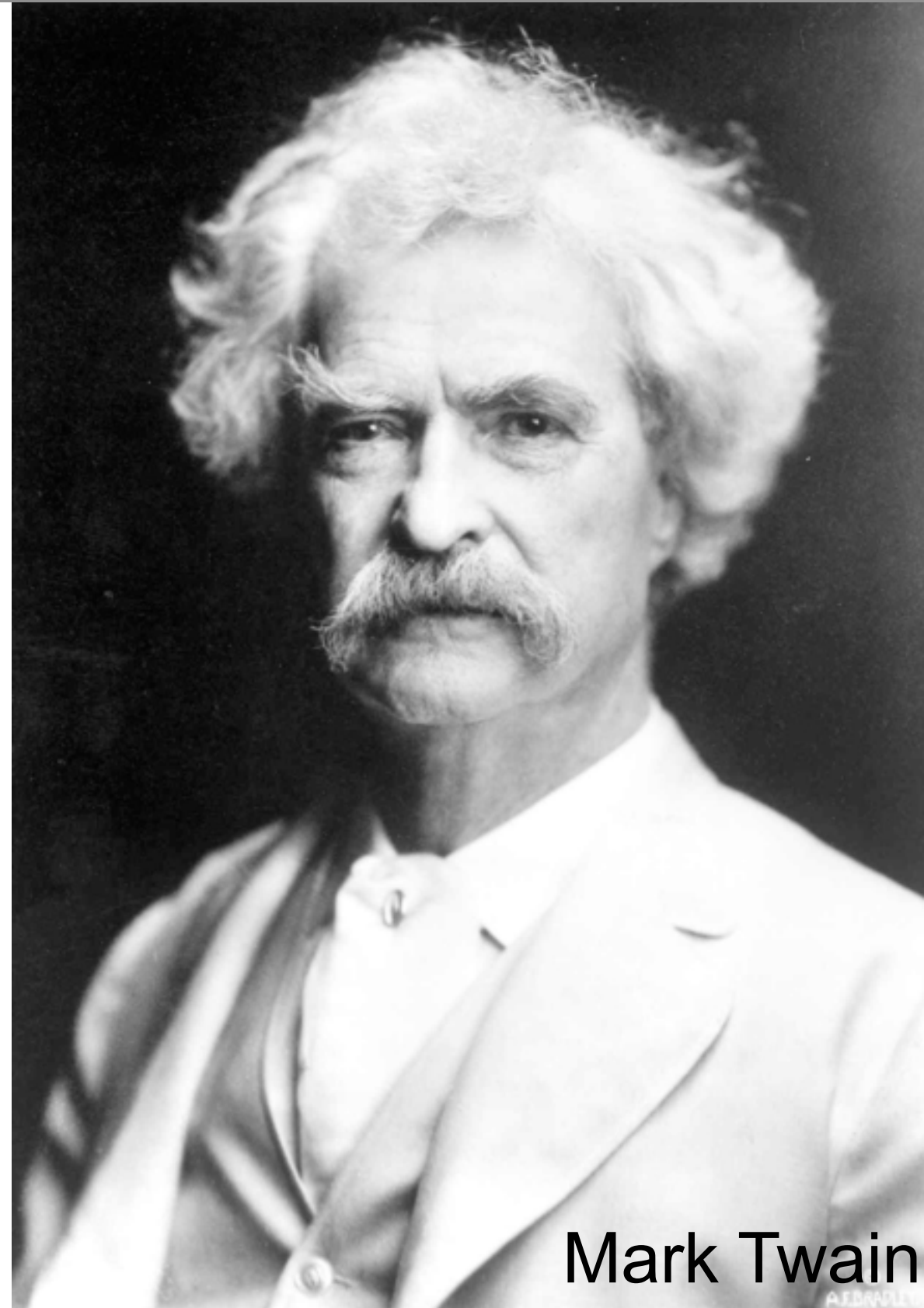
Mark Twain

The Awful German Language

"A dog is **der Hund**; now you put that dog in the genitive case, and is he the same dog he was before? No, sir; he is **des Hundes**; put him in the dative case and what is he? Why, he is **dem Hund**. Now you snatch him into the accusative case and how is it with him? Why, he is **den Hunden**. But suppose he happens to be twins and you have to pluralize him- what then? Why, they'll swat that twin dog around through the 4 cases until he'll think he's an entire international dog show. I don't like dogs, but I wouldn't treat a dog like that. "
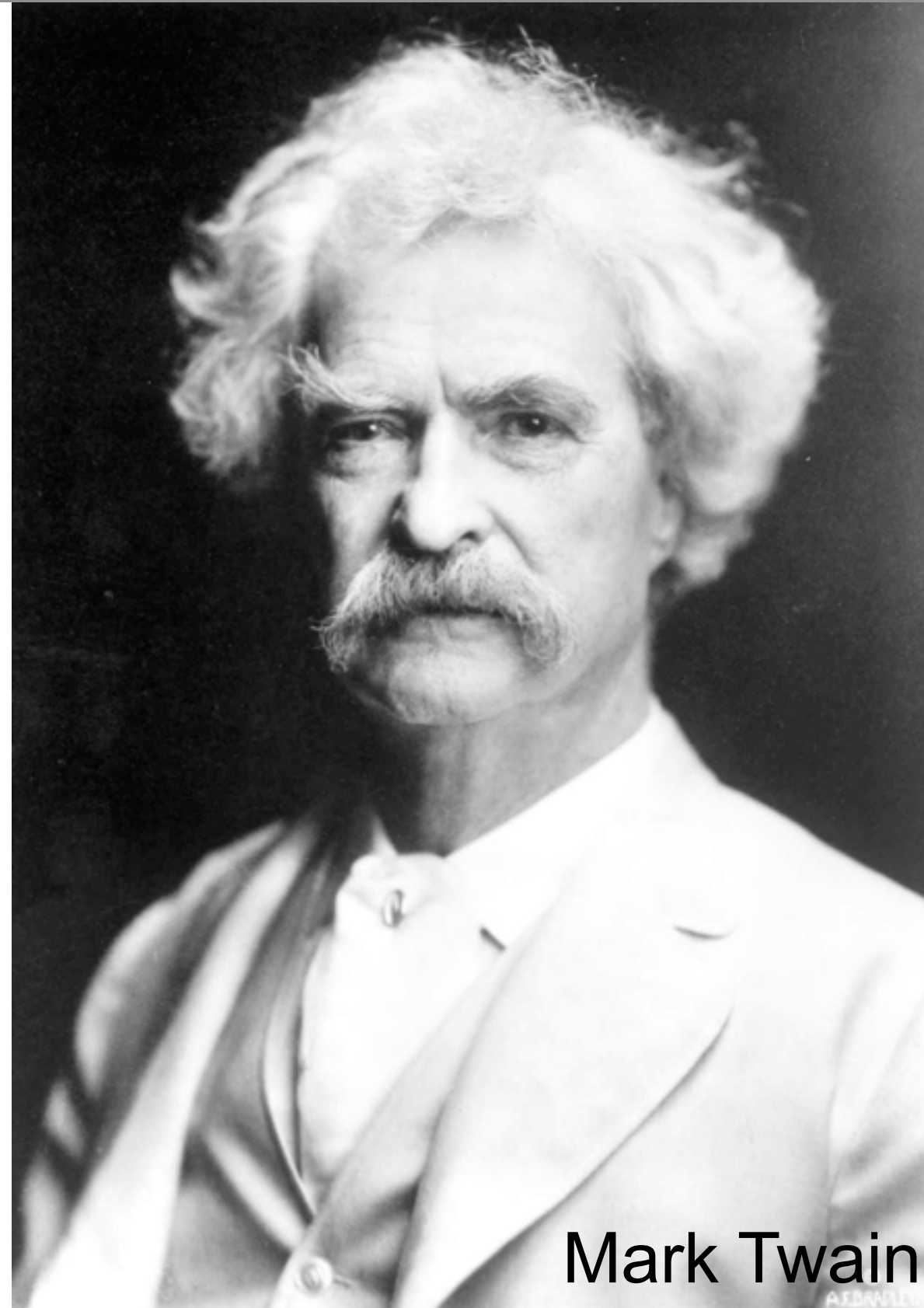
Mark Twain

The Awful German Language

"The Germans have an inhuman way of cutting up their verbs. Now a verb has a hard time enough of it in this world when it's all together. It's downright inhuman to split it up. But that's just what those Germans do. They take part of a verb and put it down here, like a stake, and they take the other part of it and put it away over yonder like another stake, and between these two limits they just shovel in German. "

Mark Twain