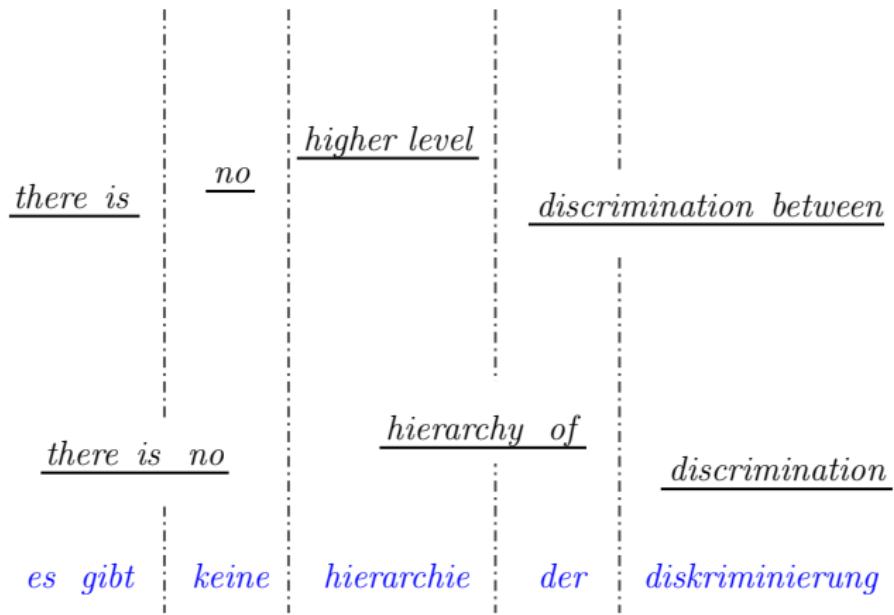


Non-Projective Parsing for Statistical Machine Translation

Xavier Carreras and Michael Collins

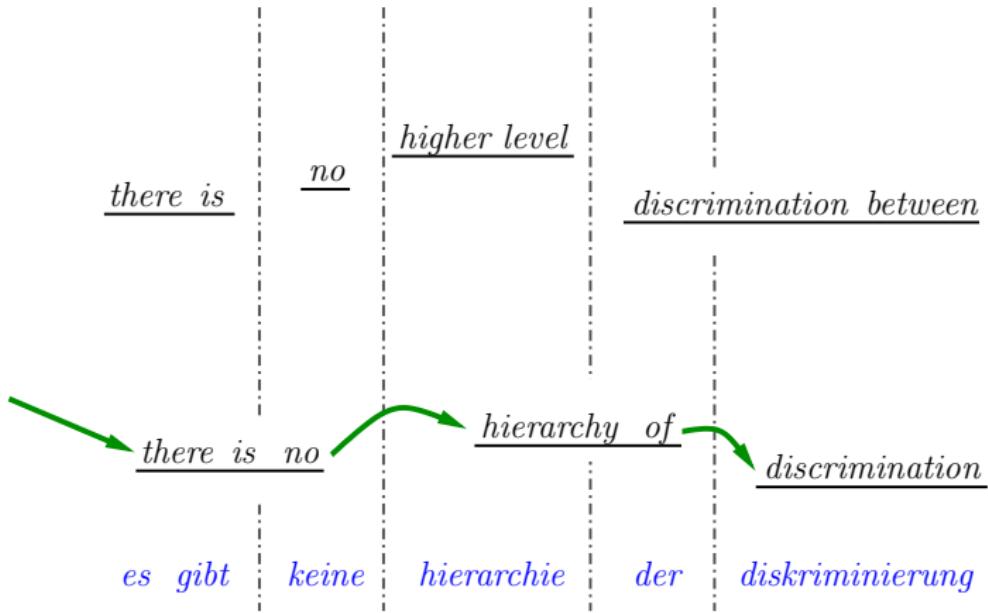
MIT CSAIL

Phrase-based Translation



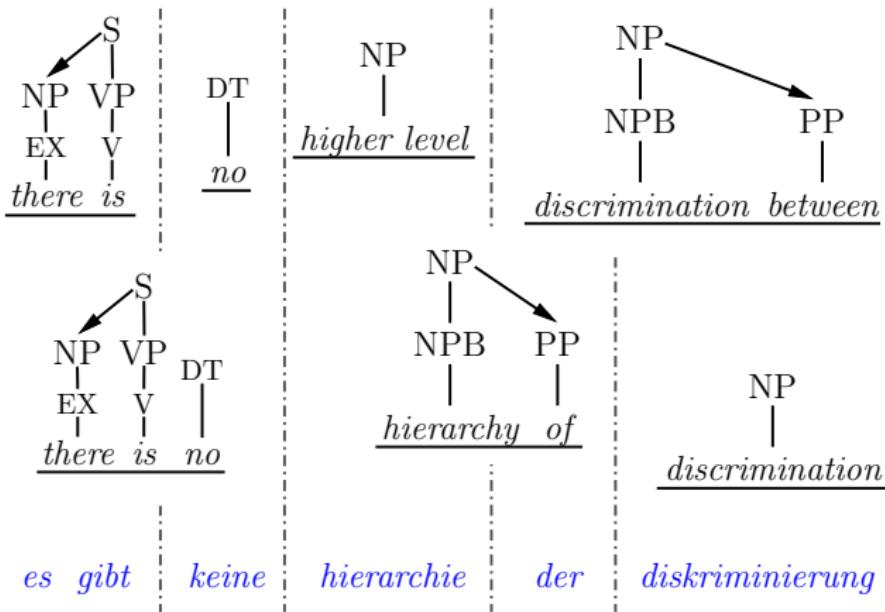
segmentation + phrase selection + distortion

Phrase-based Translation



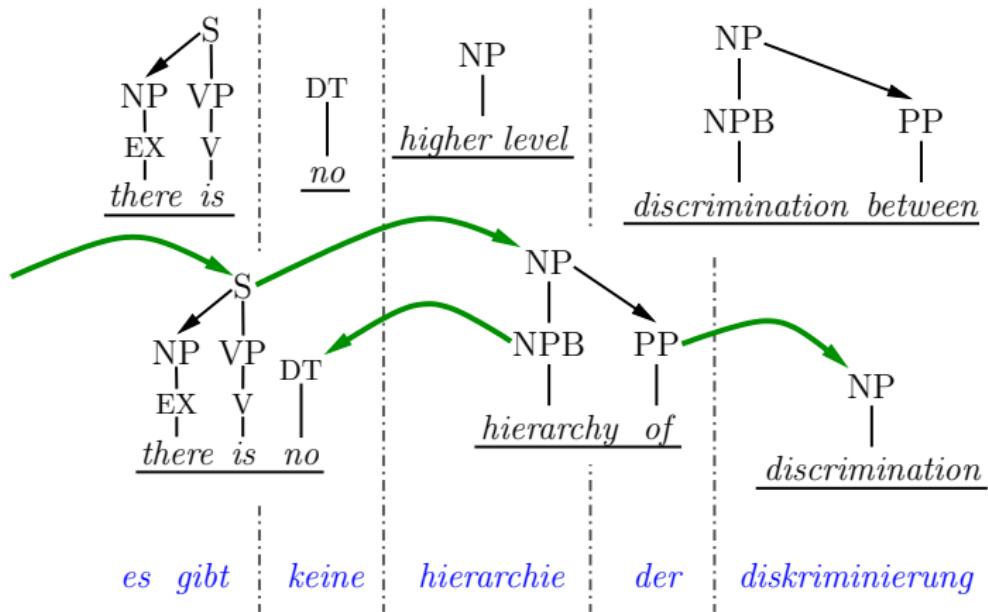
segmentation + phrase selection + distortion

Phrase-based Translation with TAG operations



segmentation + s-phrase selection + adjunctions

Phrase-based Translation with TAG operations



segmentation + s-phrase selection + adjunctions

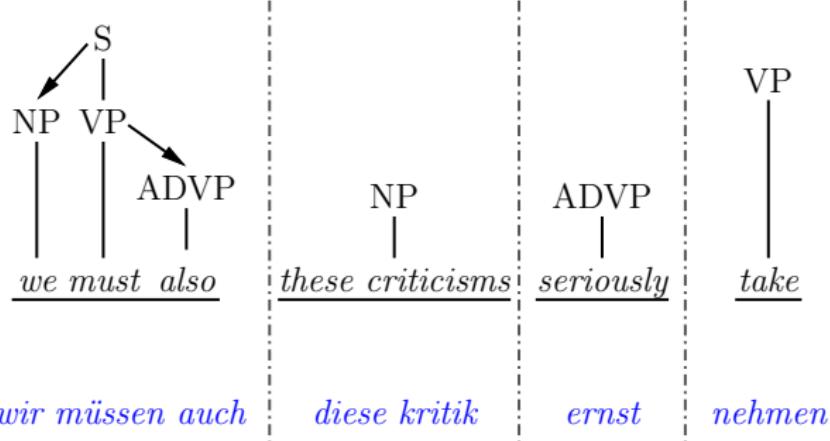
Phrase-based Translation with TAG operations

Contributions (I)

A TAG-based syntactic translation model. Properties:

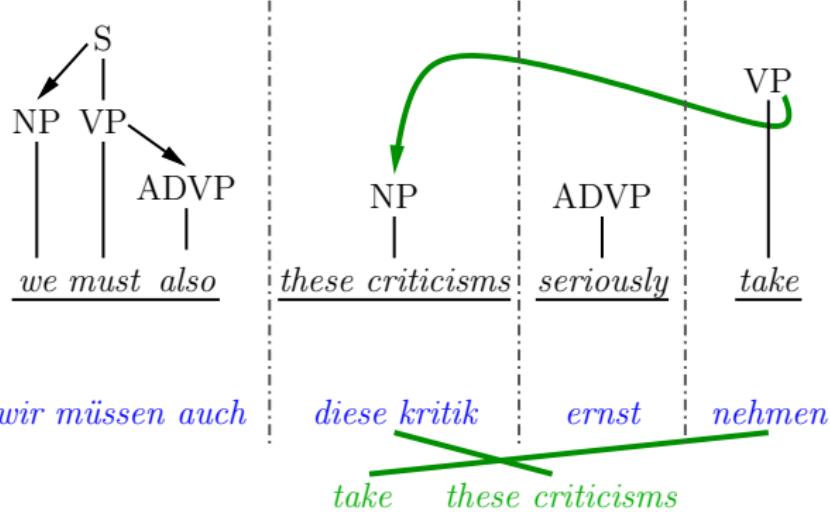
- ▶ Retains the full set of lexical entries of a phrase-based system
- ▶ Straightforward integration of a syntactic language model

Reordering via Non-Projective Operations



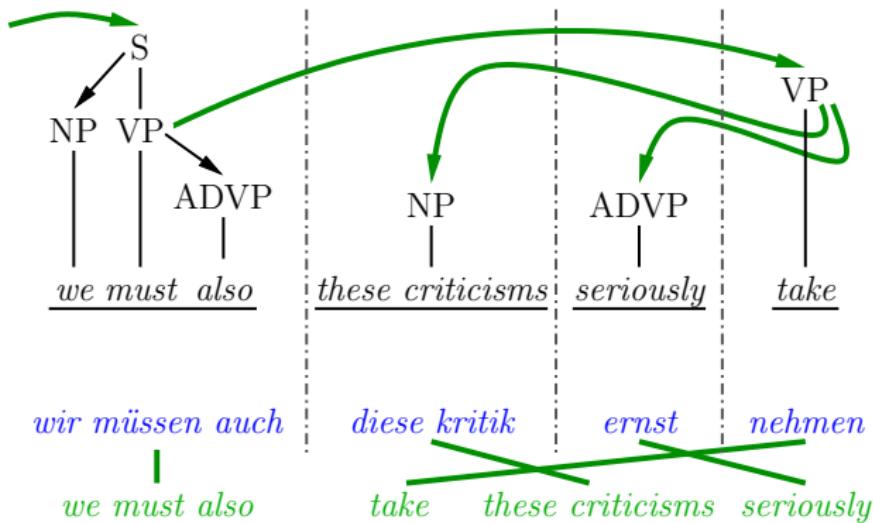
segmentation + s-phrase selection + non-projective adjunctions

Reordering via Non-Projective Operations



segmentation + s-phrase selection + non-projective adjunctions

Reordering via Non-Projective Operations



segmentation + s-phrase selection + non-projective adjunctions

Reordering via Non-Projective Operations

Contributions (II)

We model reordering with flexible non-projective adjunctions.

- ▶ How to control reorderings?
 - ▶ A discriminative model inspired by work in dependency parsing (e.g. [McDonald et al. 05])
 - ▶ Hard constraints
- ▶ How to decode efficiently?
 - ▶ A novel beam-search algorithm

Outline

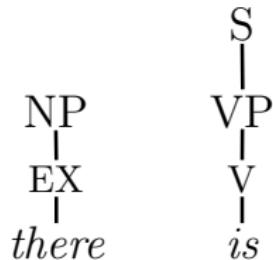
Translation as TAG-based Parsing

Constraints on Reorderings

Efficient Decoding

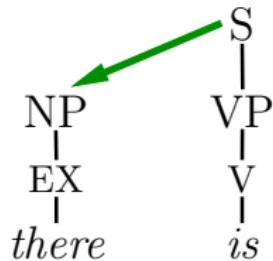
Experiments

A TAG formalism [Carreras, Collins and Koo 2008]



- ▶ Basic units are **spines**
- ▶ Spines are combined using **adjunction** operations

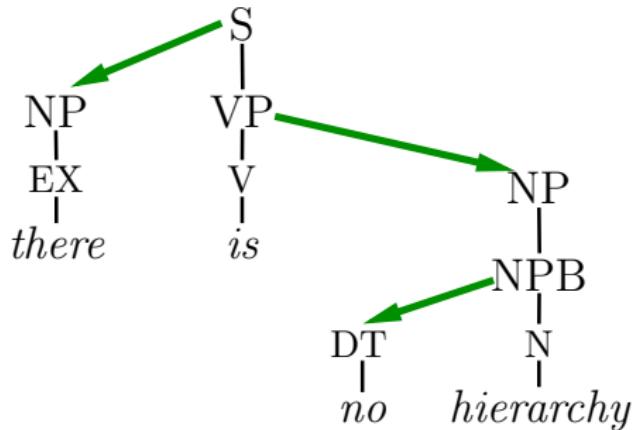
A TAG formalism [Carreras, Collins and Koo 2008]



QUESTION

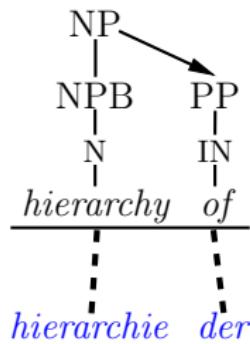
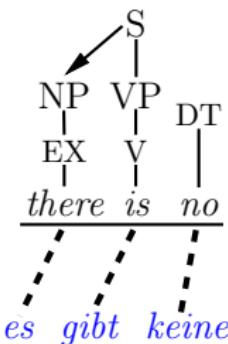
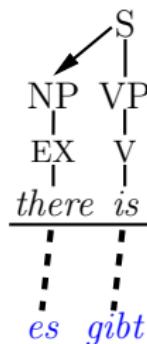
- ▶ Basic units are **spines**
- ▶ Spines are combined using **adjunction** operations

A TAG formalism [Carreras, Collins and Koo 2008]



- ▶ Basic units are **spines**
- ▶ Spines are combined using **adjunction** operations

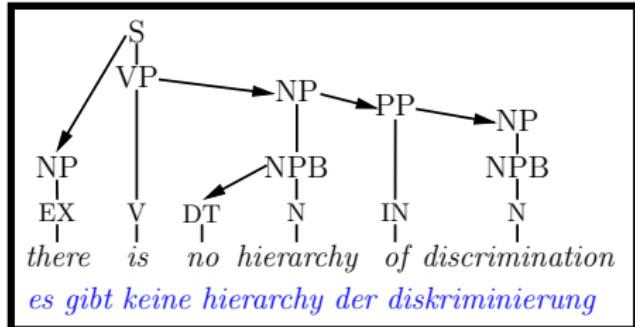
S-phrases: Syntactic Phrase-entries for Translation



An s-phrase consists of:

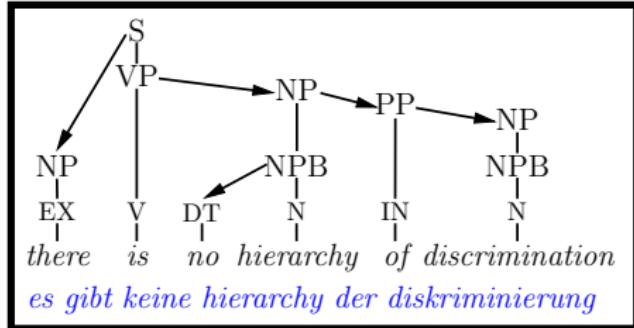
- ▶ Foreign words
- ▶ English words
- ▶ A syntactic structure
- ▶ An alignment

Extraction of S-phrases



- ▶ Training example = source sentence + English sentence + English parse tree
- ▶ We use phrasal entries from a standard phrase-based approach

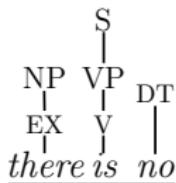
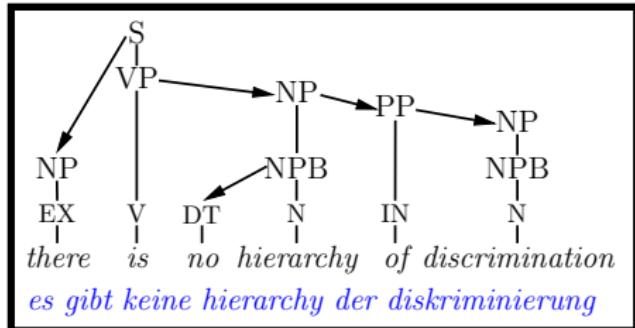
Extraction of S-phrases



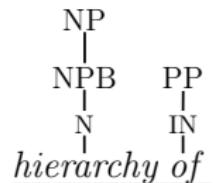
there is no hierarchy of
es gibt keine hierarchie der

- ▶ Training example = source sentence + English sentence + English parse tree
- ▶ We use phrasal entries from a standard phrase-based approach

Extraction of S-phrases



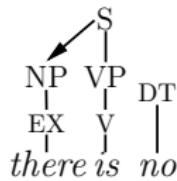
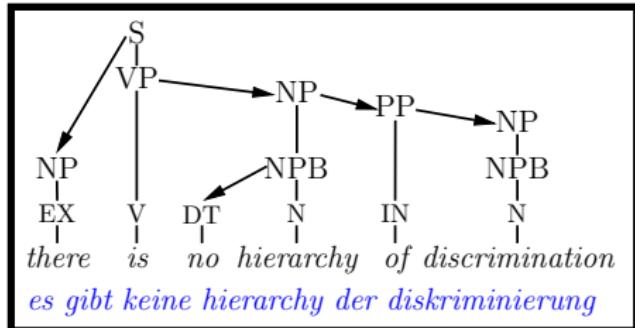
es gibt keine hierarchie der



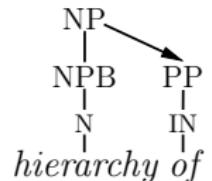
hierarchy of

- ▶ Training example = source sentence + English sentence + English parse tree
- ▶ We use phrasal entries from a standard phrase-based approach

Extraction of S-phrases



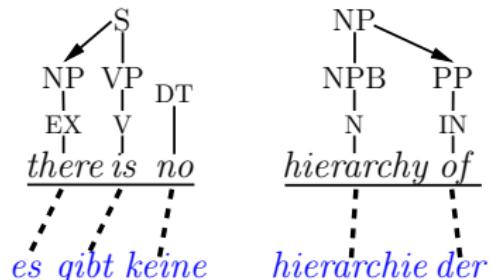
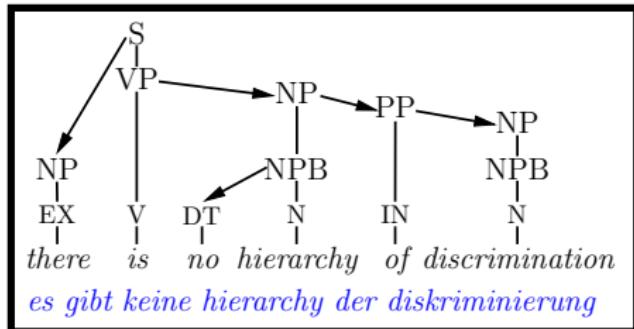
es gibt keine hierarchy of discrimination



hierarchy of discrimination

- ▶ Training example = source sentence + English sentence + English parse tree
- ▶ We use phrasal entries from a standard phrase-based approach

Extraction of S-phrases



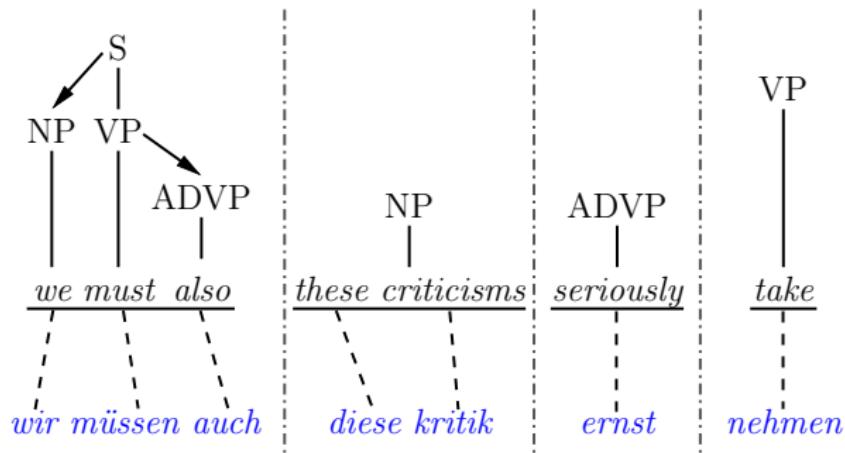
- ▶ Training example = source sentence + English sentence + English parse tree
- ▶ We use phrasal entries from a standard phrase-based approach

Derivations

wir müssen auch diese kritik ernst nehmen

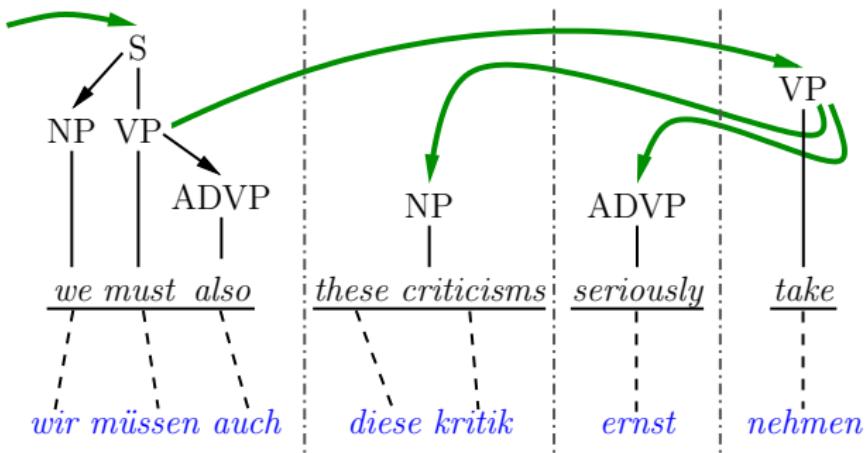
- ▶ A derivation:

Derivations



- ▶ A derivation:
 - ▶ Step 1: segment the input sentence,
and choose an s-phrase for each segment

Derivations



- ▶ A derivation:
 - ▶ Step 1: segment the input sentence, and choose an s-phrase for each segment
 - ▶ Step 2: connect s-phrases with adjunctions

Model

- ▶ Model score for a derivation d :

$$\begin{aligned} \textit{score}(d) &= \textit{score}_{LM}(d) + \textit{score}_P(d) \\ &\quad + \textit{score}_{SYN}(d) + \textit{score}_R(d) \end{aligned}$$

where

- ▶ \textit{score}_{LM} is a trigram language model
- ▶ \textit{score}_P is a sum of standard phrase-based scores
- ▶ \textit{score}_{SYN} is a syntactic language model [Charniak et al. 03]
[Shen et al. 08] (probabilities are associated with adjunctions)
- ▶ \textit{score}_R is a sum of discriminative adjunction scores

Model

- ▶ Model score for a derivation d :

$$\begin{aligned} \textit{score}(d) &= \textcolor{red}{\textit{score}_{LM}(d)} + \textit{score}_P(d) \\ &\quad + \textit{score}_{SYN}(d) + \textit{score}_R(d) \end{aligned}$$

where

- ▶ \textit{score}_{LM} is a trigram language model
- ▶ \textit{score}_P is a sum of standard phrase-based scores
- ▶ \textit{score}_{SYN} is a syntactic language model [Charniak et al. 03]
[Shen et al. 08] (probabilities are associated with adjunctions)
- ▶ \textit{score}_R is a sum of discriminative adjunction scores

Model

- ▶ Model score for a derivation d :

$$\begin{aligned} \textit{score}(d) &= \textit{score}_{LM}(d) + \textcolor{red}{\textit{score}_P(d)} \\ &\quad + \textit{score}_{SYN}(d) + \textit{score}_R(d) \end{aligned}$$

where

- ▶ \textit{score}_{LM} is a trigram language model
- ▶ $\textcolor{red}{\textit{score}_P}$ is a sum of standard phrase-based scores
- ▶ \textit{score}_{SYN} is a syntactic language model [Charniak et al. 03]
[Shen et al. 08] (probabilities are associated with adjunctions)
- ▶ \textit{score}_R is a sum of discriminative adjunction scores

Model

- ▶ Model score for a derivation d :

$$\begin{aligned} \textit{score}(d) &= \textit{score}_{LM}(d) + \textit{score}_P(d) \\ &+ \textcolor{red}{\textit{score}_{SYN}(d)} + \textit{score}_R(d) \end{aligned}$$

where

- ▶ \textit{score}_{LM} is a trigram language model
- ▶ \textit{score}_P is a sum of standard phrase-based scores
- ▶ $\textcolor{red}{\textit{score}_{SYN}}$ is a syntactic language model [Charniak et al. 03]
[Shen et al. 08] (probabilities are associated with adjunctions)
- ▶ \textit{score}_R is a sum of discriminative adjunction scores

Model

- ▶ Model score for a derivation d :

$$\begin{aligned} \textit{score}(d) &= \textit{score}_{LM}(d) + \textit{score}_P(d) \\ &+ \textit{score}_{SYN}(d) + \textcolor{red}{\textit{score}_R(d)} \end{aligned}$$

where

- ▶ \textit{score}_{LM} is a trigram language model
- ▶ \textit{score}_P is a sum of standard phrase-based scores
- ▶ \textit{score}_{SYN} is a syntactic language model [Charniak et al. 03]
[Shen et al. 08] (probabilities are associated with adjunctions)
- ▶ $\textcolor{red}{\textit{score}_R}$ is a sum of discriminative adjunction scores

Two challenges

All permutations of s-phrases are possible.

Two challenges:

1. Constraining reorderings
2. Search

Outline

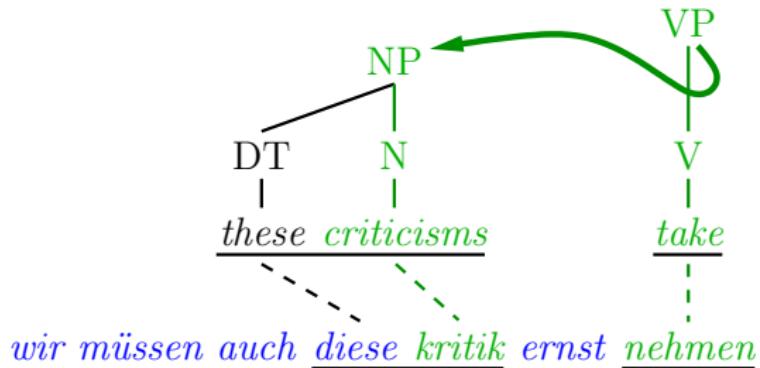
Translation as TAG-based Parsing

Constraints on Reorderings

Efficient Decoding

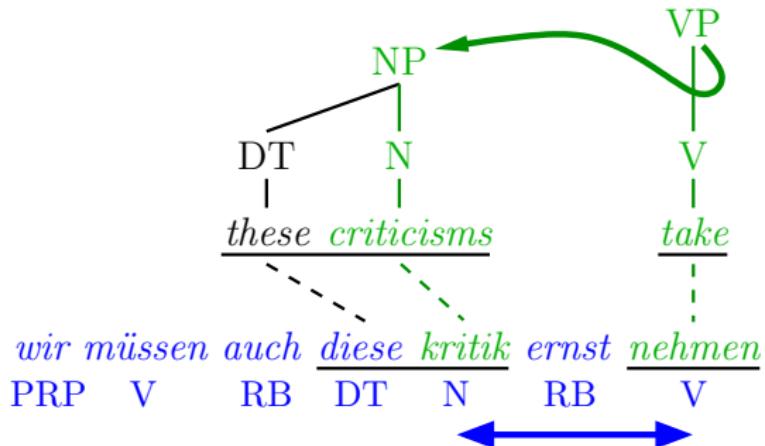
Experiments

$score_R$: A Discriminative Dependency Model



$score_R(d)$ is a **discriminative dependency model** (related to work in dependency parsing (e.g. [McDonald et al. 05]))

$score_R$: A Discriminative Dependency Model



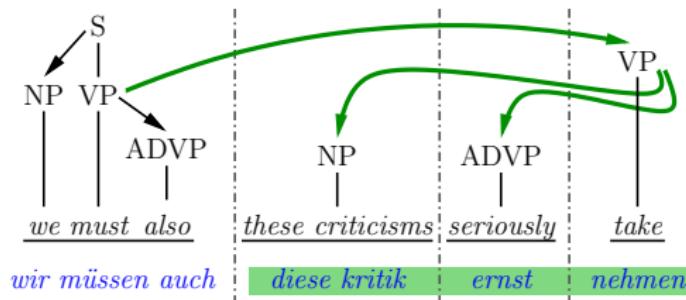
$score_R(d)$ is a **discriminative dependency model** (related to work in dependency parsing (e.g. [McDonald et al. 05]))

π -constituent constraint

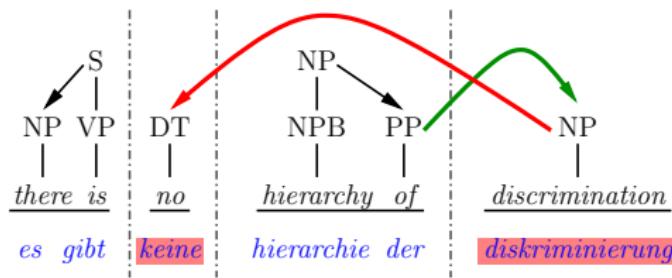
Define π -constituent: a head spine with all its descendants

Constraint any π -constituent must be aligned to a contiguous substring in the source sentence

Satisfied:



Violated:



Outline

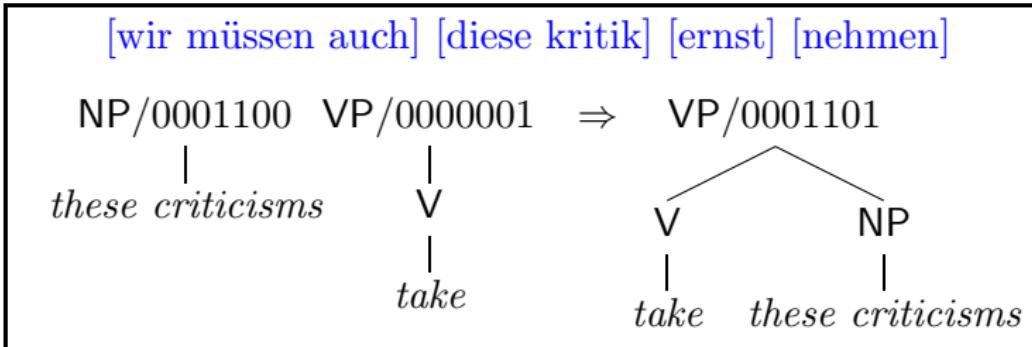
Translation as TAG-based Parsing

Constraints on Reorderings

Efficient Decoding

Experiments

Decoding as Parsing



- ▶ Projective parsing: each constituent has an associated **span**
- ▶ A generalization: each constituent has a **bit-string** recording which foreign words have been translated
- ▶ Beam search strategy: ensures that the top N analyses for each foreign word are explored at each stage

Outline

Translation as TAG-based Parsing

Constraints on Reorderings

Efficient Decoding

Experiments

Experiments

German to English using Europarl data (750K training sentences)

Development:

System	BLEU score
Syntax-based	25.2
Syntax (no disc. model)	23.7 (-1.5)
Syntax (no π -c constraint)	24.4 (-0.8)

Experiments

German to English using Europarl data (750K training sentences)

Development:

System	BLEU score
Syntax-based	25.2
Syntax (no disc. model)	23.7 (-1.5)
Syntax (no π -c constraint)	24.4 (-0.8)

Test:

System	BLEU score
Phrase-based system (Pharaoh)	24.58
Syntax-based system	25.04 (+0.46)

significant ($p = 0.021$) under paired bootstrap resampling [Koehn 04]
close to significant ($p = 0.058$) under the sign test [Collins et al. 05]

Human Evaluations

Ref: Now, however, we are seeing that president Putin is pursuing a policy of openness towards the west.

Now, however, we see that mr president Putin is pursuing a policy of openness towards the west.

We are, however, now that president Putin a policy of openness to the west out of blackmail.

Human Evaluations

Ref: Now, however, we are seeing that president Putin is pursuing a policy of openness towards the west.

Syn: Now, however, we see that mr president Putin is pursuing a policy of openness towards the west.

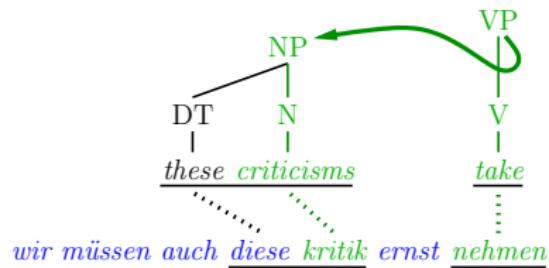
PB : We are, however, now that president Putin a policy of openness to the west out of blackmail.

	Syntax	PB	=	Total
Syntax	51	3	7	61
PB	1	25	11	37
=	21	14	67	102
Total	73	42	85	200

both results are significant with $p < 0.05$ under the sign test

Summary

A TAG-based syntactic translation model

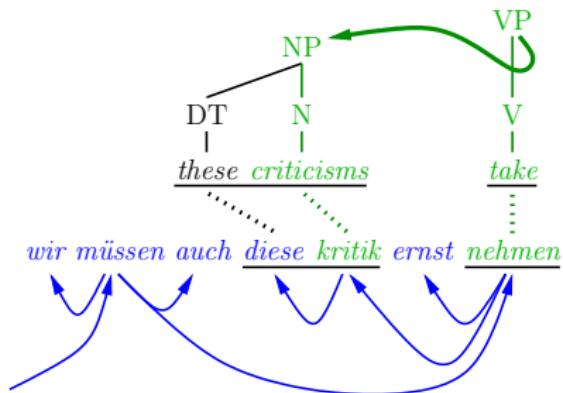


Non-projective adjunctions for reordering:

- ▶ Arbitrary reorderings
- ▶ Discriminative dependency model

Summary

A TAG-based syntactic translation model



Non-projective adjunctions for reordering:

- ▶ Arbitrary reorderings
- ▶ Discriminative dependency model

Future work: Condition on syntactic structure of the source string