A Discriminative Model for Tree-to-Tree Translation

Michael Collins, MIT

(Joint work with Brooke Cowan, Ivona Kučerová)

Statistical Machine Translation

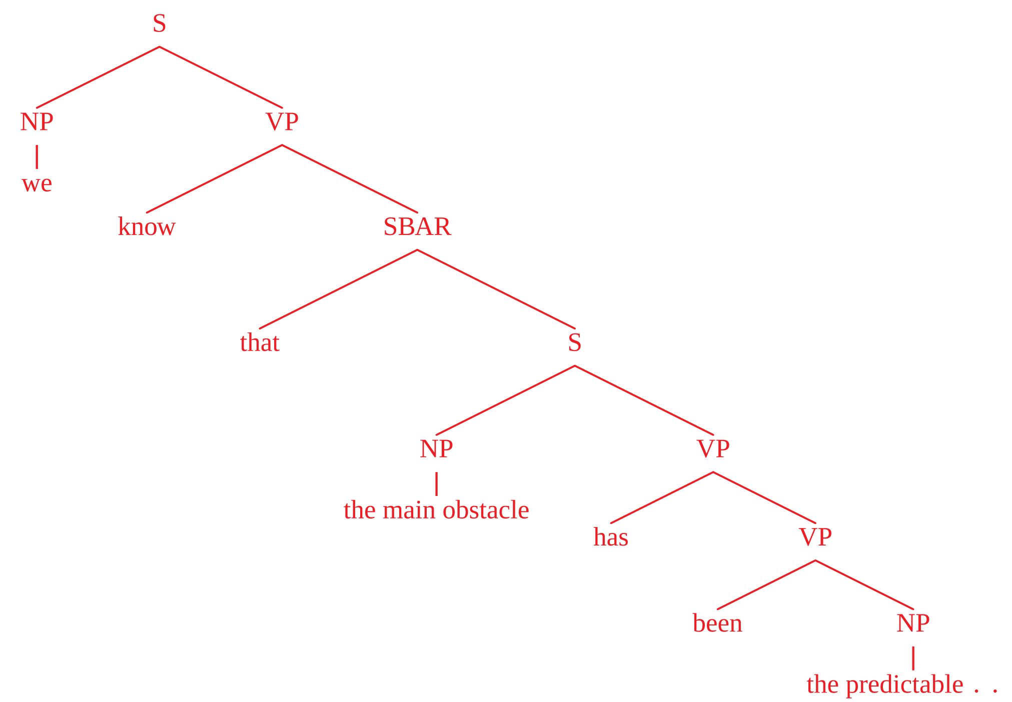
- Idea due to IBM, in early 1990's: given many examples of French-English translations, can we automatically learn a translation model?

- **Data**: Canadian Parliamentary Proceedings (French-English); European parliament (translations amongst 12 European languages); United Nations data (Arabic-English, Chinese-English).

Our Goal: Tree-to-Tree Translation

s np-sb wir
  vafin-hd wissen
  s-oc kous-cp daß
    np-sb art das
      nn haupthemmnis
    np-pd art der
      adja vorhersehbare
      nn widerstand
      np-ag art der
        nn hersteller
    vafin-hd war

$\Rightarrow$



- Training data: parse both sides of a bitext with existing parsers
- E.g., for Europarl, we have around 750,000 (x, y) pairs where each x is a German parse, each y is an English parse

Why Syntax?

- Predicting a *target language tree* allows us to enforce syntactic constraints on the output

- Use of the *source language tree* allows us to take into account grammatical relations in the source language, and reflect these in the target language tree

Phrase-Based Systems

Step 1: Automatic Learning of a Phrasal Lexicon:

Today we shall be debating the reopening of the Mont Blanc tunnel

Heute werden wir uber die Wiedereroffnung des Mont-Blanc-Tunnels diskutieren.

$\Downarrow$

Today	$\leftrightarrow$	Heute
we shall be	$\leftrightarrow$	werden wir
the reopening	$\leftrightarrow$	die Wiedereroffnung
of the Mont Blanc tunnel	$\leftrightarrow$	des Mont-Blanc-Tunnels
debating	$\leftrightarrow$	diskutieren

Given training data of $\approx$ 300,000 sentence pairs, induce close to a million phrase pairs, with associated frequencies

Phrase-Based Systems

Step 2 Translate using a greedy, left-to-right decoding method

Today

Heute werden wir uber die Wiedereroffnung des Mont-Blanc-Tunnels diskutieren.

$$\text{Score} = \underbrace{\log P(\text{Today} \mid \text{START})}_{\text{Language model}}$$

$$+ \underbrace{\log P(\text{Heute} \mid \text{Today})}_{\text{Phrase model}}$$

$$+ \underbrace{\log P(\text{1-1} \mid \text{1-1})}_{\text{Distortion model}}$$

Phrase-Based Systems

Step 3 Translate using a greedy, left-to-right decoding method

Today [we shall be]

Heute [werden wir] uber die Wiedereroffnung des Mont-Blanc-Tunnels diskutieren.

$$\text{Score} = \underbrace{\log P(\text{we shall be} \mid \text{today})}_{\text{Language model}}$$

$$+ \underbrace{\log P(\text{werden wir} \mid \text{we will be})}_{\text{Phrase model}}$$

$$+ \underbrace{\log P(\text{2-3} \mid \text{3-4})}_{\text{Distortion model}}$$

Phrase-Based Systems

Step 3 Translate using a greedy, left-to-right decoding method

Today we shall be debating

Heute werden wir uber die Wiedereroffnung des Mont-Blanc-Tunnels diskutieren.

Phrase-Based Systems

Step 3 Translate using a greedy, left-to-right decoding method

Today we shall be debating the reopening

Heute werden wir uber die Wiedereroffnung des Mont-Blanc-Tunnels diskutieren.

Phrase-Based Systems

Step 3 Translate using a greedy, left-to-right decoding method

Today we shall be debating the reopening of the Mont Blanc tunnel

Heute werden wir uber die Wiedereroffnung

des Mont-Blanc-Tunnels diskutieren.

Phrase-Based Systems

Step 3 Translate using a greedy, left-to-right decoding method

Today we shall be debating the reopening of the Mont Blanc tunnel

Heute werden wir uber die Wiedereroffnung des Mont-Blanc-Tunnels diskutieren.

Why Syntax?

- Output from phrase-based systems:

Reference: consequently proposals are submitted to parliament under the assent procedure, meaning that parliament can no longer table amendments, as directives in this area were adopted as single market legislation under the codecision procedure on the basis of art.100a tec.

Translation: consequently, the proposals parliament after the assent procedure, the tabled amendments for offers no possibility of community directives, because as part of the internal market legislation on the basis of article 100a of the treaty in the codecision procedure have been adopted.

- Intuition/working hypothesis:
  - phrase-based systems are very good at predicting *content words*,
  - but are less accurate in producing *function words*, or producing output that correctly encodes *grammatical relations between content words*

Why Syntax?

Previous work (Collins, Koehn and Kučerová, ACL 2005):

s np-adv **yesterday**
vafin-hd **has**
np-sb **he**
v-inf **said**
s-oc kous-cp **that**
np-sb art **the**
nn **main obstacle**
np-pd art **the**
adja **predictable**
nn **resistance**
np-ag art **of the**
nn **manufacturers**
vafin-hd was

⇒

s np-adv **yesterday**
np-sb **he**
vafin-hd **has**
v-inf **said**
s-oc kous-cp **that**
np-sb art **the**
nn **main obstacle**
vafin-hd was
np-pd art **the**
adja **predictable**
nn **resistance**
np-ag art **of the**
nn **manufacturers**

- Source language reordered in training and test data
- BLEU scores increased from 25.2 to 26.8
  Human evaluations: 40 translations improved, 40 same, 20 worse

Related Work on Syntax in MT

- Synchronous grammars: Wu, 1995; Alshawi, 1996; Melamed, 2004; Zhang, Gildea, 2006; Synchronous TAG: Shieber and Schabes, 1990; Nesson, Shieber, Rush, 2006

- Dependency formalisms: Quirk, Menezes, Cherry, 2005; Ding, Palmer, 2005. LFG: Riezler, Maxwell, 2006

- Syntactic MT at ISI: Yamada and Knight, 2001; Graehl and Knight, 2004; Galley, Hopkins, Knight, and Marcu, 2004; Galley et. al, 2006; Marcu, Wang, Echihabi, Knight, 2006

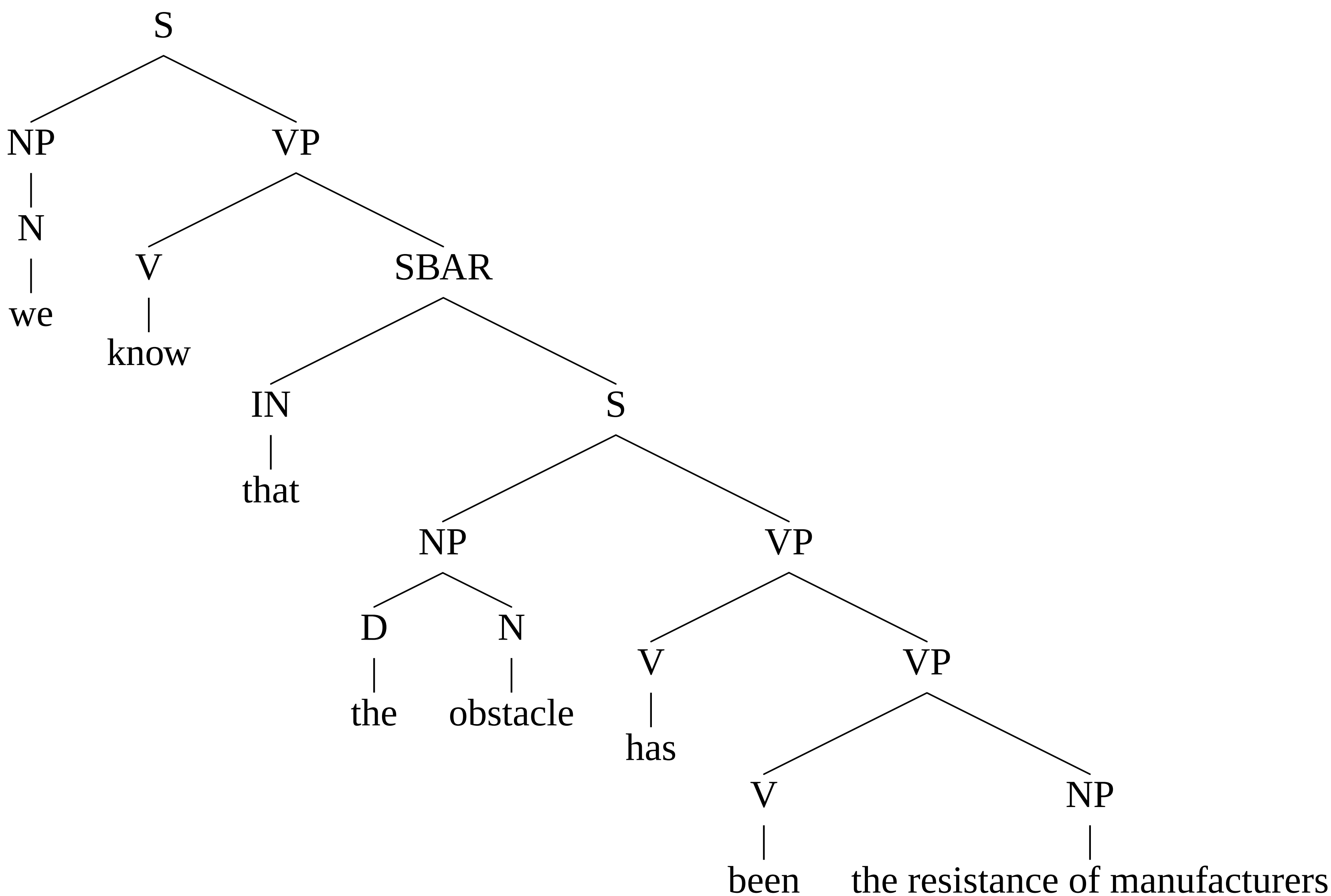
- Hiero (Chiang 2005)

- Discriminative approaches: Turian, Wellington, Melamed, 2006; Wellington, Turian Melamed, 2006

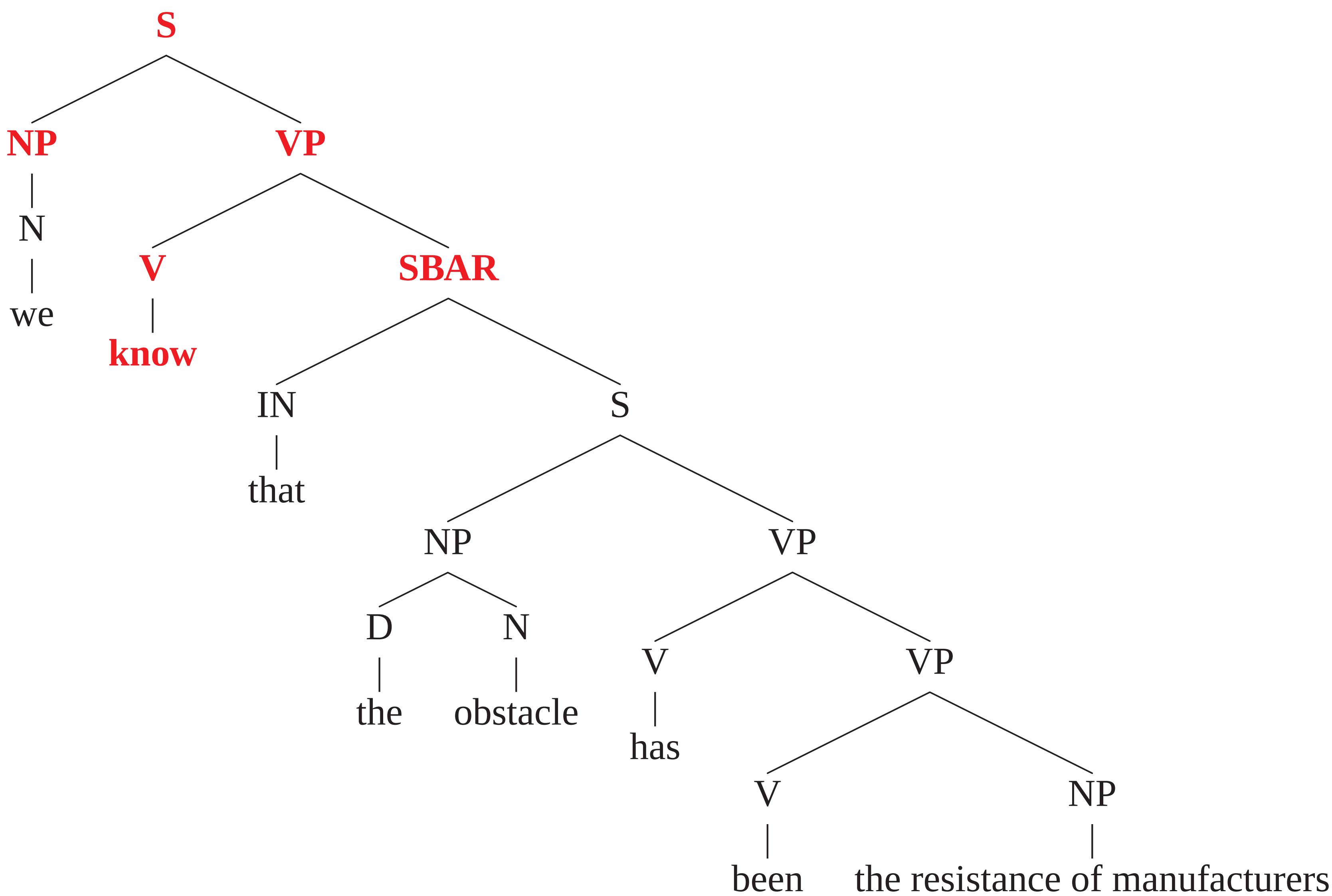
Outline

- An architecture for tree-to-tree translation
- Training the model using the perceptron algorithm
- Experiments
- Conclusions

Extended Projections (EPs) (Grimshaw, 1991; Frank, 2002)

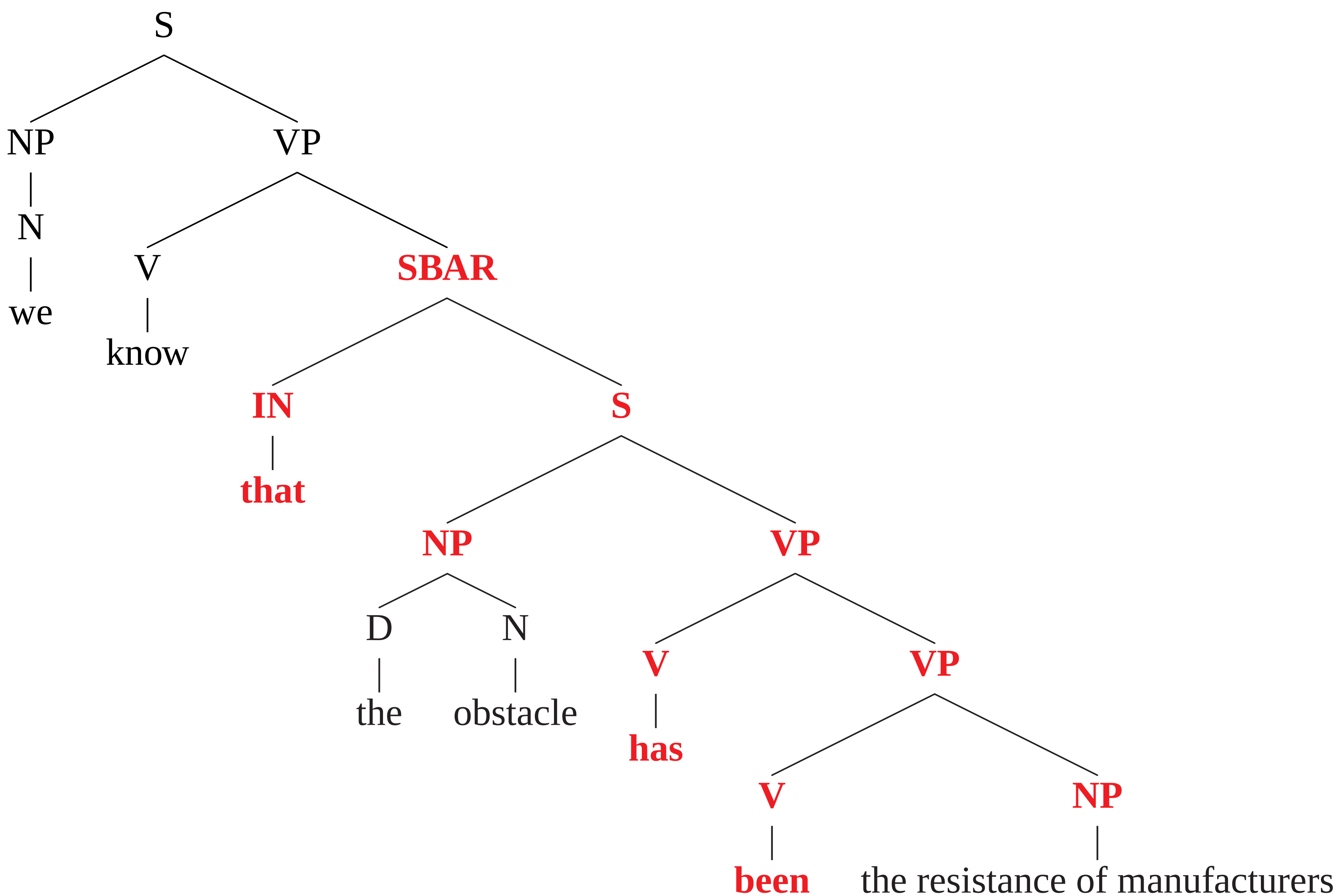


Extended Projections (EPs) (Grimshaw, 1991; Frank 2002)



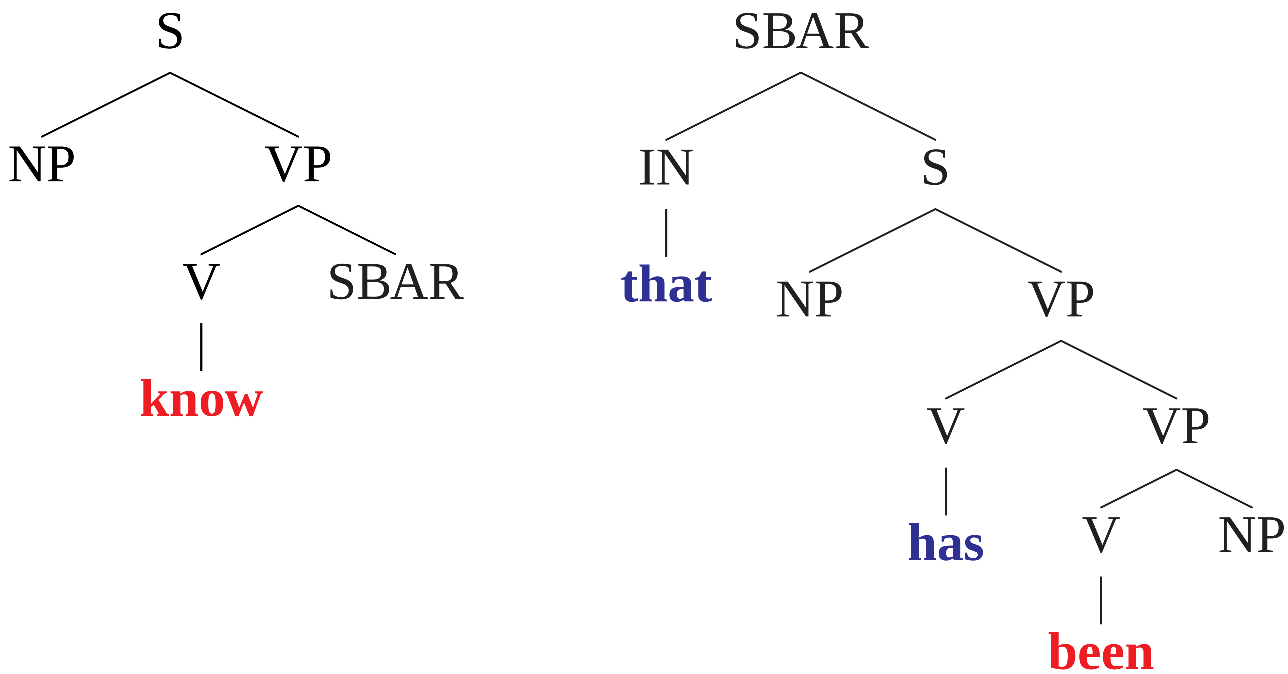
EP for *know*

Extended Projections (EPs) (Grimshaw, 1991; Frank, 2002)



EP for *been*

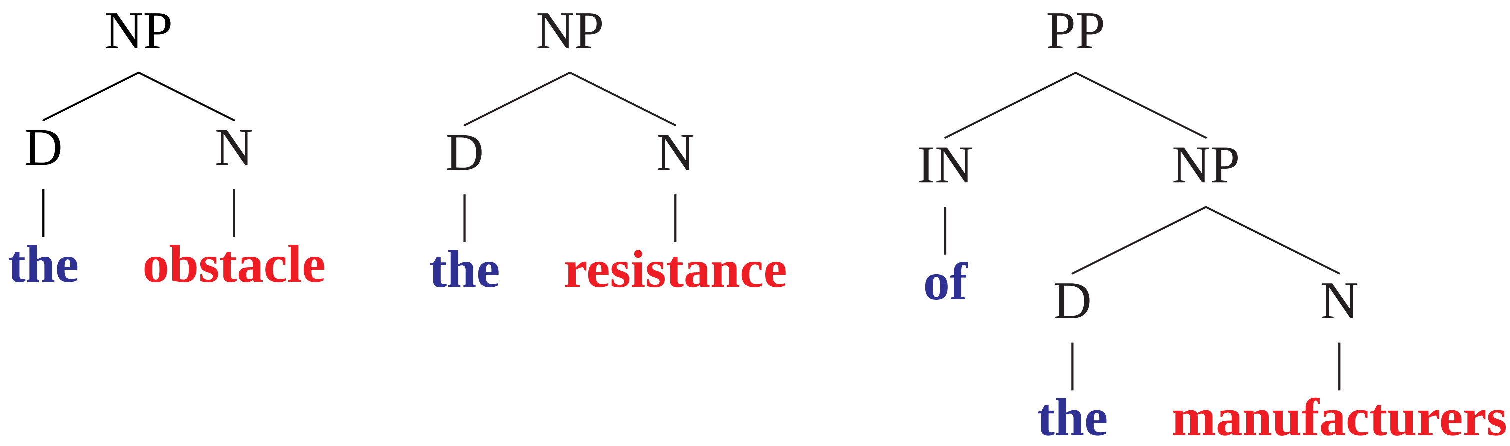
Some Examples of Extended Projections



Each extended projection consists of:

- A content word (e.g., know, been)
- A sub-tree
- 0 or more function words

Some More Examples of Extended Projections



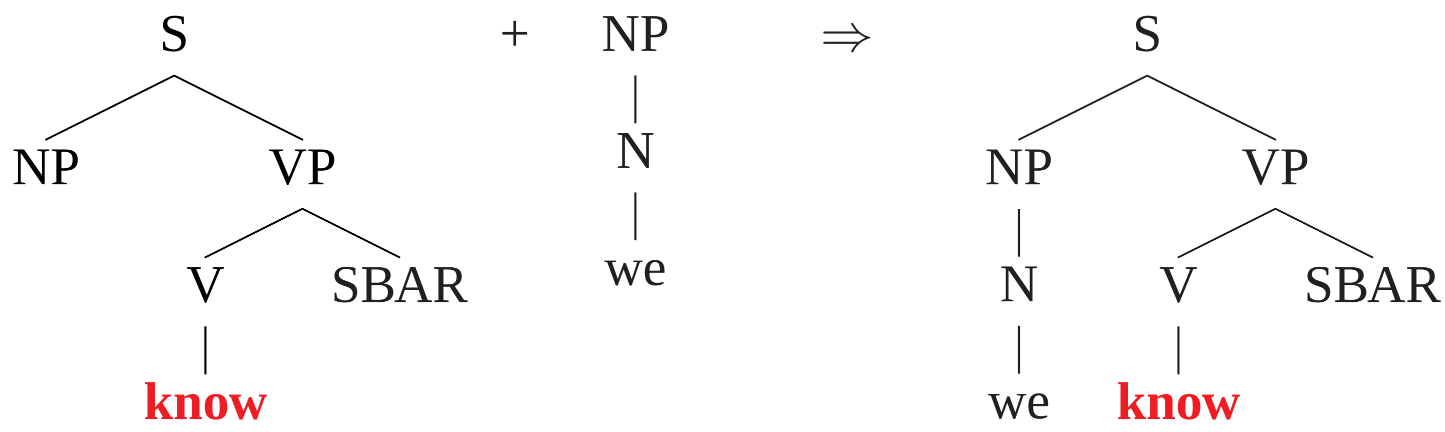
- Each content word (noun, main verb, adjective, adverb) has an extended projection

- **We'll concentrate on EPs for verbs**

Extended Projections in Tree Adjoining Grammar (Frank, 2002)

- Each EP is a Tree-adjoining grammar (TAG) elementary tree

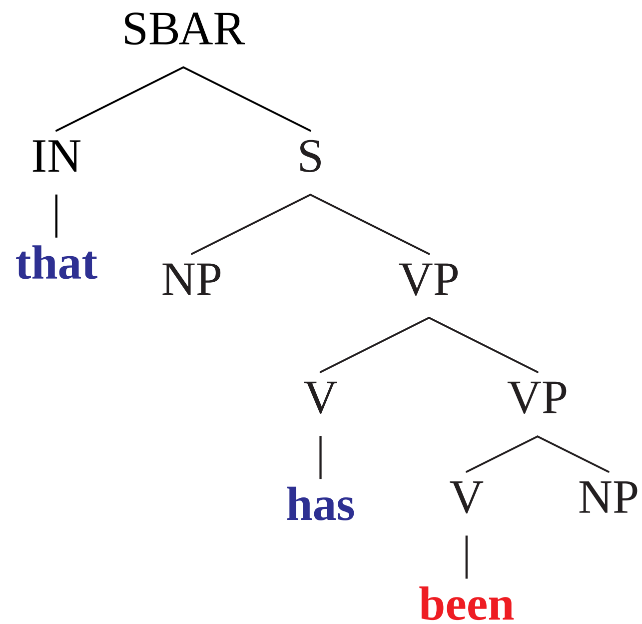
- EP's can be combined using TAG operations, e.g.,



Predicting Extended Projections for Verbs

s-oc kous-cp daß
np-sb[1] art das
nn haupthemmnis
np-pd[2] art der
adja vorhersehbare
nn widerstand
np-ag art der
nn hersteller
vafin-hd war

⇒



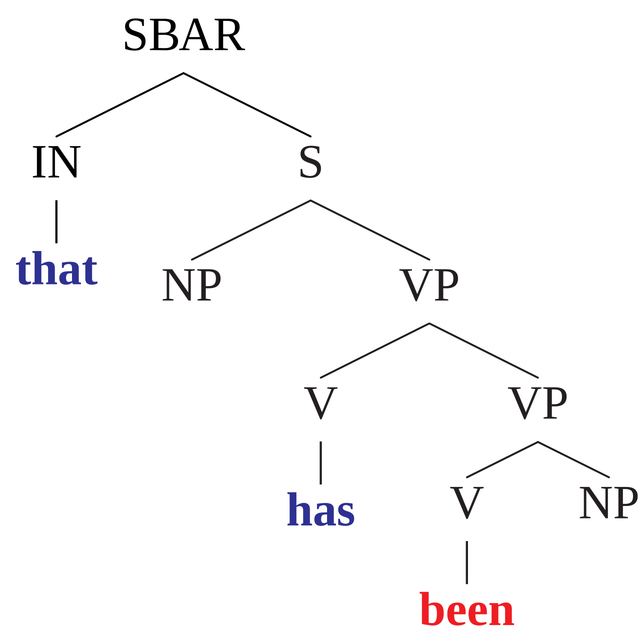
Paraphrase:

that
[np-sb[1] the main obstacle]
[np-pd[2] the predictable resistance of manufacturers]
was

Adding Alignment Information

```
s-oc kous-cp daß
     np-sb[1] art das
              nn haupthemmnis
     np-pd[2] art der
              adja vorhersehbare
              nn widerstand
              np-ag art der
                    nn hersteller
     vafin-hd war
```

$\Rightarrow$

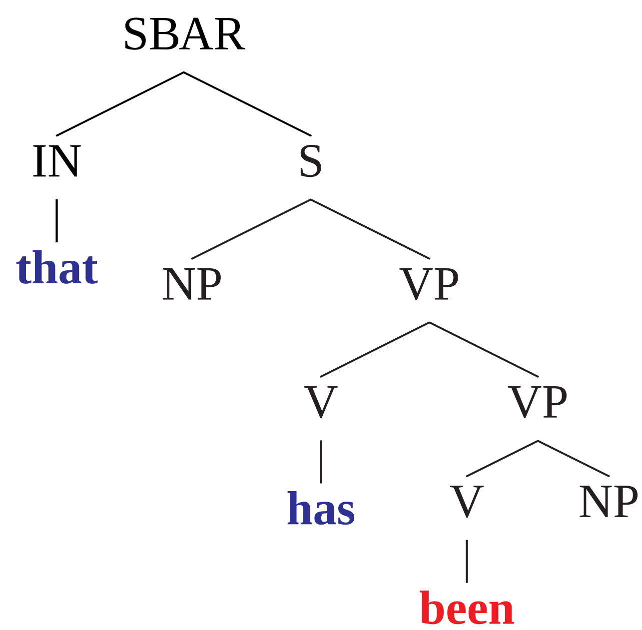


- In this case there are two German modifiers, [1] and [2]
- Alignment information specifies a *target position* for each German modifier
- We call an Extended Projection + alignment information an **Aligned Extended Projection (AEP)**

An Example AEP

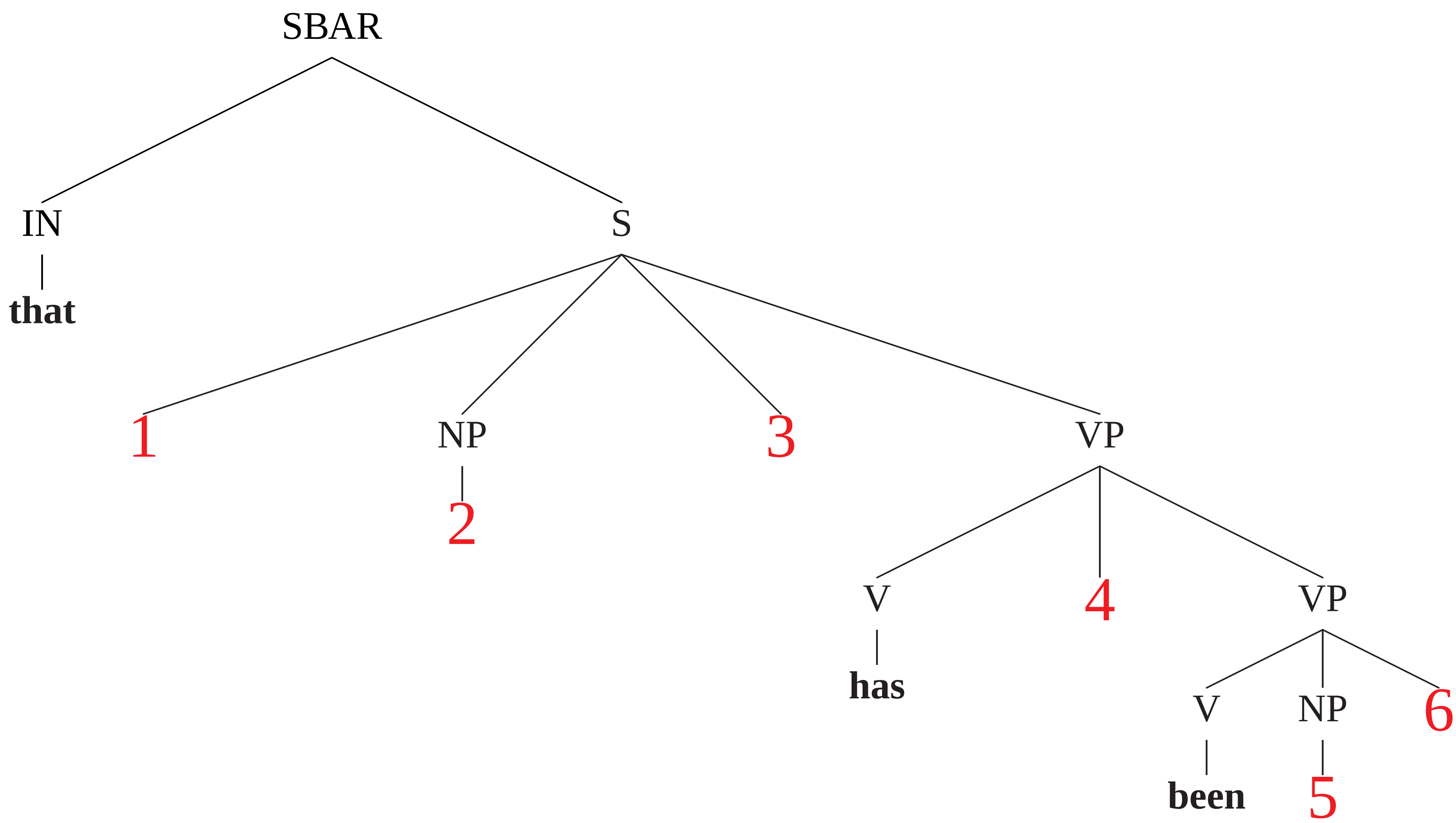
```
s-oc kous-cp daß
     np-sb[1] art das
              nn haupthemmnis
     np-pd[2] art der
              adja vorhersehbare
              nn widerstand
              np-ag art der
                    nn hersteller
     vafin-hd war
```

$\Rightarrow$



[1] = SUBJECT
[2] = OBJECT

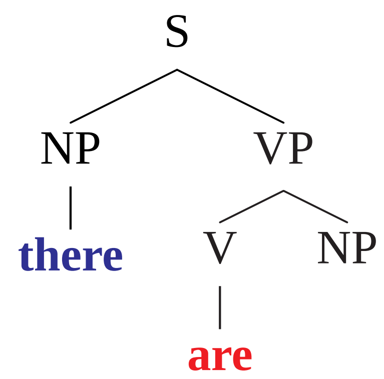
Possible Modifier Positions



An Example AEP

s pp-mo[1] appr zwischen
piat beiden
nn gesetzen
vvfin-hd bestehen
adv-mo[2] also
np-sb[3] adja erhebliche
adja rechtliche
$, ,
adja praktische
kon und
adja wirtschaftliche
nn unterschiede

$\Rightarrow$



[1] = POST-VERB
[2] = PRE-SUBJECT
[3] = OBJECT

Paraphrase:
[pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]

Final translation:
so there are significant legal, practical and economic differences between the two pieces of legislation

Translation Using AEPs

Step 1: Break German parse tree into a sequence of clauses

[S wir wissen [S-OC daß das haupthemmnis der vorsehbare . . . war]]

⇓

s np-sb $\boxed{1}$ wir
  vafin-hd wissen

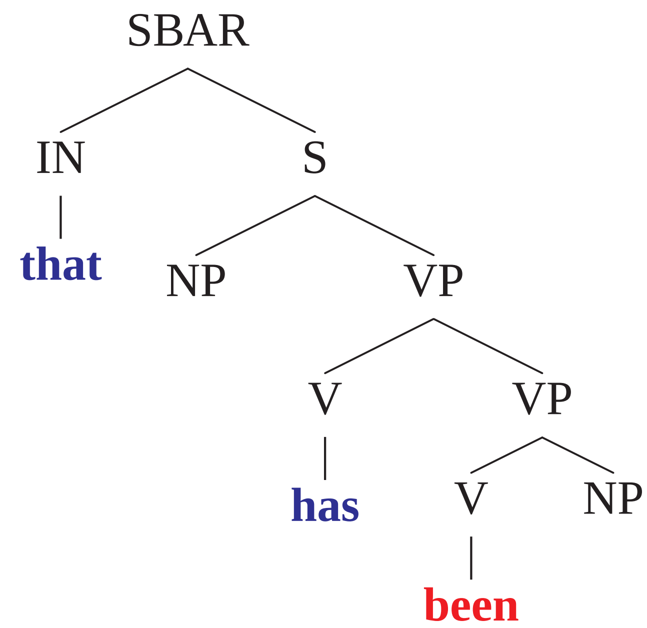
s-oc kous-cp daß
  np-sb $\boxed{1}$ art das
    nn haupthemmnis
  np-pd $\boxed{2}$ art der
    adja vorhersehbare
    nn widerstand
    np-ag art der
      nn hersteller
  vafin-hd war

Translation Using AEPs

Step 2: For each German clause, predict an AEP

s-oc kous-cp daß
  np-sb[1] art das
    nn haupthemmnis
  np-pd[2] art der
    adja vorhersehbare
    nn widerstand
    np-ag art der
      nn hersteller
  vafin-hd war

⇒



[1] = SUBJECT
[2] = OBJECT

Translation Using AEPs

Step 3: Translate each modifier using a phrase-based system, and insert modifiers in the AEP

das haupthemmnis	$\Rightarrow$	the main obstacle
der vorhersehbare widerstand . . .	$\Rightarrow$	the predictable resistance . . .

Final translation:
that the main obstacle has been the predictable resistance . . .

Translation Using AEPs

Step 4: Combine the English clauses

we know

$+$

that the main obstacle has been the predictable resistance . . .

$\Downarrow$

we know that the main obstacle has been the predictable resistance . . .

Motivation

- Clause-by-clause translation considerably simplifies the tree-to-tree translation problem, and seems a reasonable assumption

Motivation

- Clause-by-clause translation considerably simplifies the tree-to-tree translation problem, and seems a reasonable assumption

- For now, our focus is on AEPs at the *clause level*:

  – Clausal structures are particularly rich from a syntactic perspective: 1) There should be a lot of potential for improved translation quality; 2) These should be a good test-bed for the methods we use (which could be applied to other structures, e.g., NPs)

Motivation

- Clause-by-clause translation considerably simplifies the tree-to-tree translation problem, and seems a reasonable assumption

- For now, our focus is on AEPs at the *clause level*:

  – Clausal structures are particularly rich from a syntactic perspective: 1) There should be a lot of potential for improved translation quality; 2) These should be a good test-bed for the methods we use (which could be applied to other structures, e.g., NPs)

- Step 3: Modifier translation with a phrase-based system

  – Modifiers are small chunks (NPs, PPs, ADJPs, etc.) which are handled well by phrase-based systems

Outline

- An architecture for tree-to-tree translation
- A Perceptron model for AEP prediction
- Experiments
- Conclusions

Extracting AEPs from a Corpus

- Input: an English sentence/parse-tree paired with a German sentence/parse-tree
- Output: 0 or more $\langle$German clause, English AEP$\rangle$ pairs
- Motivation: gather training data for a model that maps German clauses to English AEPs
- Method: algorithm that makes use of parse trees $+$ alignments

Note: high precision, low recall, but still gives over 450,000 training examples for 750,000 sentences of Europarl data

Representing AEPs as Decision Sequences

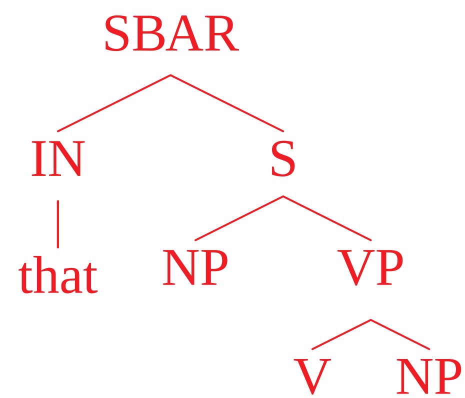
- Our training set is (x_i, y_i) for $i = 1 \ldots m$, where each x_i is a German clause, each y_i is an English AEP.

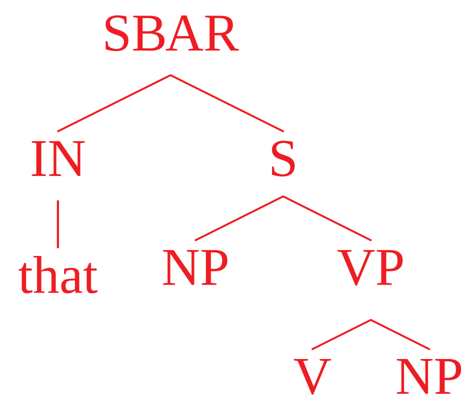
- We represent each y_i as the sequence of decisions

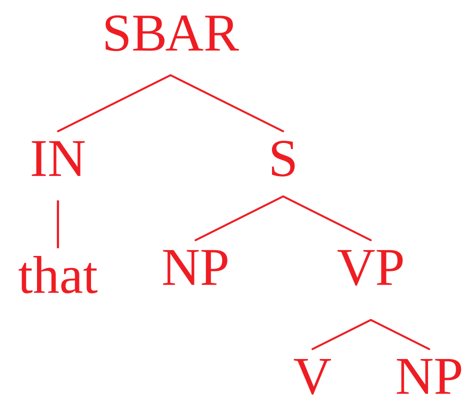
  $\langle$`STEM, SPINE, VOICE, SUBJECT, OBJECT, WH, MODALS, INFL, MOD(1), MOD(2), ..., MOD(n)`$\rangle$

German Clause	English AEP
s-oc kous-cp daß np-sb[1] art das nn haupthemmnis np-pd[2] art der adja vorhersehbare nn widerstand np-ag art der nn hersteller vafin-hd war Paraphrase: *that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was*	STEM: SPINE: VOICE: SUBJECT: OBJECT: WH: MODALS: INFL: MOD1: MOD2:

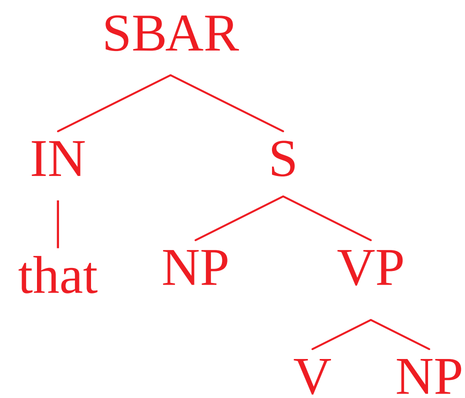
German Clause	English AEP
s-oc kous-cp daß np-sb$\boxed{1}$ art das nn haupthemmnis np-pd$\boxed{2}$ art der adja vorhersehbare nn widerstand np-ag art der nn hersteller vafin-hd war Paraphrase: *that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was*	STEM: be SPINE: VOICE: SUBJECT: OBJECT: WH: MODALS: INFL: MOD1: MOD2:

German Clause	English AEP
s-oc kous-cp daß np-sb[1] art das nn haupthemmnis np-pd[2] art der adja vorhersehbare nn widerstand np-ag art der nn hersteller vafin-hd war Paraphrase: *that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was*	STEM: be SPINE:  VOICE: SUBJECT: OBJECT: WH: MODALS: INFL: MOD1: MOD2:

German Clause	English AEP
s-oc kous-cp daß np-sb[1] art das nn haupthemmnis np-pd[2] art der adja vorhersehbare nn widerstand np-ag art der nn hersteller vafin-hd war Paraphrase: *that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was*	STEM: be SPINE:  VOICE: active SUBJECT: OBJECT: WH: MODALS: INFL: MOD1: MOD2:

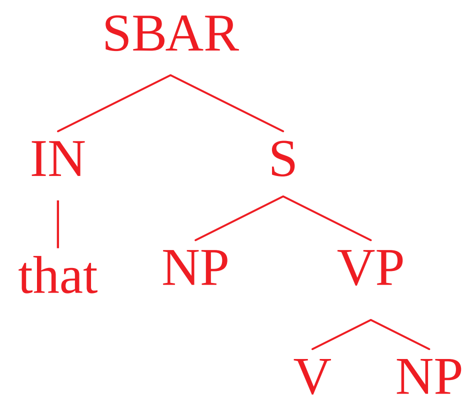
German Clause	English AEP
s-oc kous-cp daß np-sb [1] art das nn haupthemmnis np-pd [2] art der adja vorhersehbare nn widerstand np-ag art der nn hersteller vafin-hd war Paraphrase: *that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was*	STEM: be SPINE:  VOICE: active SUBJECT: [1] OBJECT: WH: MODALS: INFL: MOD1: MOD2:

German Clause	English AEP

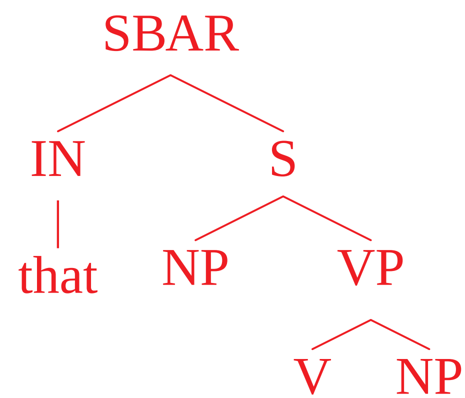
s-oc kous-cp daß
    np-sb[1] art das
        nn haupthemmnis
    np-pd[2] art der
        adja vorhersehbare
        nn widerstand
        np-ag art der
            nn hersteller
    vafin-hd war
 	
Paraphrase: *that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was*
	STEM: be
	
	SPINE: 
	
	VOICE: active
	SUBJECT: [1]
	OBJECT: [2]
	WH:
	MODALS:
	INFL:
	MOD1:
	MOD2:

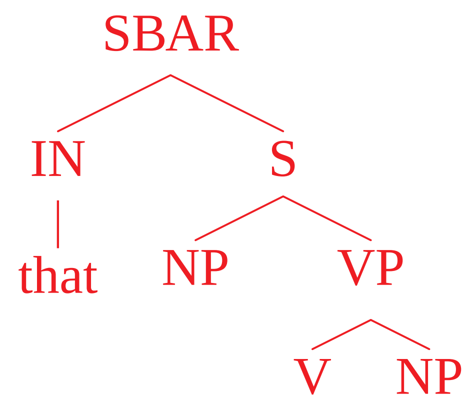
<table>
<tr><th>German Clause</th><th>English AEP</th></tr>
<tr><td>s-oc kous-cp daß
np-sb[1] art das
nn haupthemmnis
np-pd[2] art der
adja vorhersehbare
nn widerstand
np-ag art der
nn hersteller
vafin-hd war

Paraphrase: that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was</td><td>STEM: be

SPINE: 

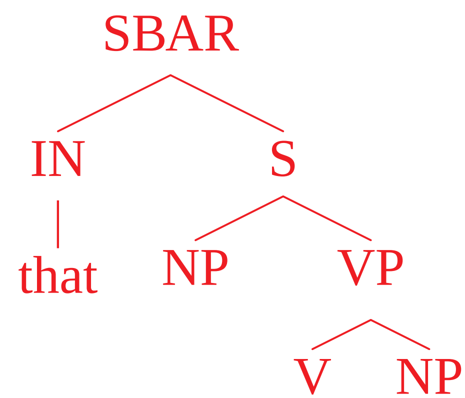
VOICE: active
SUBJECT: [1]
OBJECT: [2]
WH: NULL
MODALS:
INFL:
MOD1:
MOD2:</td></tr>
</table>

German Clause	English AEP
s-oc kous-cp daß np-sb [1] art das nn haupthemmnis np-pd [2] art der adja vorhersehbare nn widerstand np-ag art der nn hersteller vafin-hd war Paraphrase: *that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was*	STEM: be SPINE:  VOICE: active SUBJECT: [1] OBJECT: [2] WH: NULL MODALS: has INFL: MOD1: MOD2:

German Clause	English AEP
s-oc kous-cp daß np-sb [1] art das nn haupthemmnis np-pd [2] art der adja vorhersehbare nn widerstand np-ag art der nn hersteller vafin-hd war Paraphrase: *that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was*	STEM: be SPINE:  VOICE: active SUBJECT: [1] OBJECT: [2] WH: NULL MODALS: has INFL: been MOD1: MOD2:

<table>
<tr><th>German Clause</th><th>English AEP</th></tr>
<tr>
<td>s-oc kous-cp daß
np-sb[1] art das
nn haupthemmnis
np-pd[2] art der
adja vorhersehbare
nn widerstand
np-ag art der
nn hersteller
vafin-hd war

Paraphrase: that [np-sb the main obstacle] [np-pd the predictable resistance of manufacturers] was</td>
<td>STEM: be

SPINE:


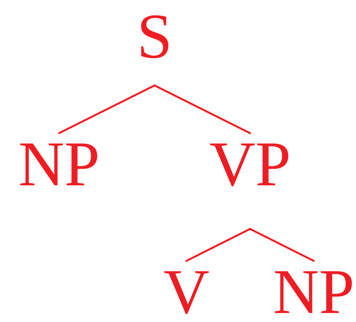
VOICE: active
SUBJECT: [1]
OBJECT: [2]
WH: NULL
MODALS: has
INFL: been
MOD1: null
MOD2: null</td>
</tr>
</table>

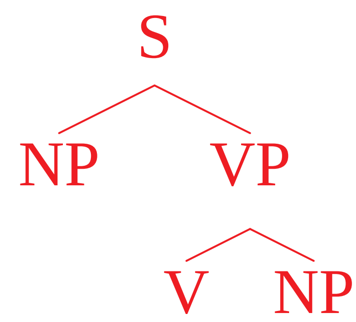
<table>
<tr><th>German Clause</th><th>English AEP</th></tr>
<tr><td>s pp-mo[1] appr zwischen
piat beiden
nn gesetzen
vvfin-hd bestehen
adv-mo[2] also
np-sb[3] adja erhebliche
adja rechtliche
$, ,
adja praktische
kon und
adja wirtschaftliche
nn unterschiede

Paraphrase: [pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]</td><td>STEM: be

SPINE:

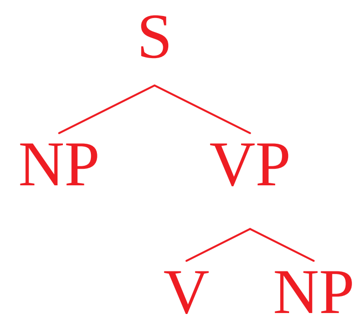
VOICE:
SUBJECT:
OBJECT:
WH:
MODALS:
INFL:
MOD1:
MOD2:
MOD3:</td></tr>
</table>

German Clause	English AEP
s pp-mo [1] appr zwischen piat beiden nn gesetzen vvfin-hd bestehen adv-mo [2] also np-sb [3] adja erhebliche adja rechtliche $, , adja praktische kon und adja wirtschaftliche nn unterschiede Paraphrase: *[pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]*	STEM: be SPINE:  VOICE: SUBJECT: OBJECT: WH: MODALS: INFL: MOD1: MOD2: MOD3:

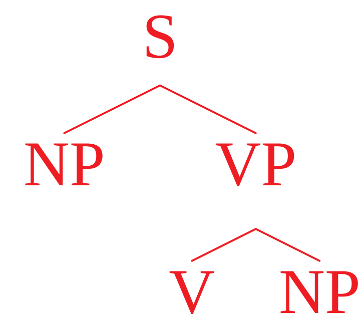
German Clause	English AEP
s pp-mo[1] appr zwischen piat beiden nn gesetzen vvfin-hd bestehen adv-mo[2] also np-sb[3] adja erhebliche adja rechtliche $, , adja praktische kon und adja wirtschaftliche nn unterschiede Paraphrase: *[pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]*	STEM: be SPINE:  VOICE: active SUBJECT: OBJECT: WH: MODALS: INFL: MOD1: MOD2: MOD3:

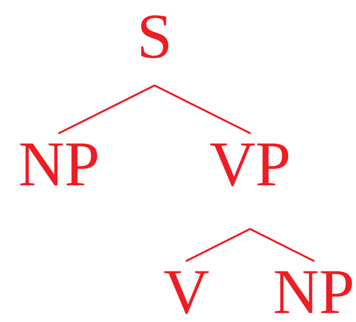
<table>
<tr><th>German Clause</th><th>English AEP</th></tr>
<tr><td>
s pp-mo[1] appr zwischen
piat beiden
nn gesetzen
vvfin-hd bestehen
adv-mo[2] also
np-sb[3] adja erhebliche
adja rechtliche
$, ,
adja praktische
kon und
adja wirtschaftliche
nn unterschiede

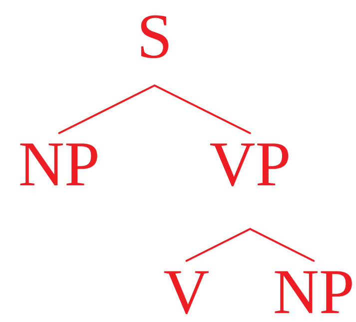
Paraphrase: [pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]
</td><td>
STEM: be

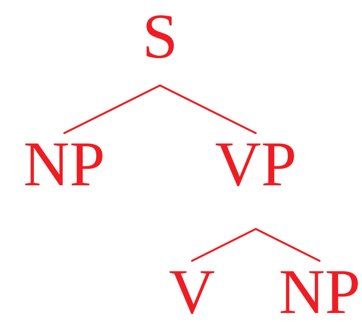
SPINE: 

VOICE: active
SUBJECT: there
OBJECT:
WH:
MODALS:
INFL:
MOD1:
MOD2:
MOD3:
</td></tr>
</table>

German Clause	English AEP
s pp-mo[1] appr zwischen piat beiden nn gesetzen vvfin-hd bestehen adv-mo[2] also np-sb[3] adja erhebliche adja rechtliche $, , adja praktische kon und adja wirtschaftliche nn unterschiede Paraphrase: *[pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]*	STEM: be SPINE:  VOICE: active SUBJECT: *there* OBJECT: [3] WH: MODALS: INFL: MOD1: MOD2: MOD3:

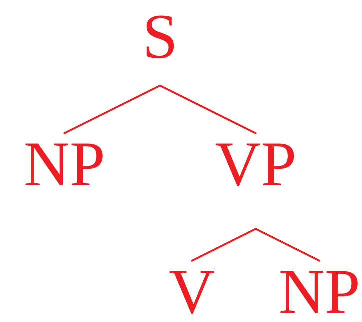
German Clause	English AEP
s pp-mo[1] appr zwischen piat beiden nn gesetzen vvfin-hd bestehen adv-mo[2] also np-sb[3] adja erhebliche adja rechtliche $, , adja praktische kon und adja wirtschaftliche nn unterschiede Paraphrase: *[pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]*	STEM: be SPINE:  VOICE: active SUBJECT: *there* OBJECT: [3] WH: NULL MODALS: INFL: MOD1: MOD2: MOD3:

German Clause	English AEP
s pp-mo $\boxed{1}$ appr zwischen piat beiden nn gesetzen vvfin-hd bestehen adv-mo $\boxed{2}$ also np-sb $\boxed{3}$ adja erhebliche adja rechtliche $, , adja praktische kon und adja wirtschaftliche nn unterschiede Paraphrase: *[pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]*	STEM: be SPINE:  VOICE: active SUBJECT: *there* OBJECT: $\boxed{3}$ WH: NULL MODALS: NULL INFL: MOD1: MOD2: MOD3:

German Clause	English AEP
s pp-mo[1] appr zwischen piat beiden nn gesetzen vvfin-hd bestehen adv-mo[2] also np-sb[3] adja erhebliche adja rechtliche $, , adja praktische kon und adja wirtschaftliche nn unterschiede Paraphrase: *[pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]*	STEM: be SPINE:  VOICE: active SUBJECT: *there* OBJECT: [3] WH: NULL MODALS: NULL INFL: are MOD1: MOD2: MOD3:

<table>
<tr><th>German Clause</th><th>English AEP</th></tr>
<tr><td>s pp-mo[1] appr zwischen
piat beiden
nn gesetzen
vvfin-hd bestehen
adv-mo[2] also
np-sb[3] adja erhebliche
adja rechtliche
$, ,
adja praktische
kon und
adja wirtschaftliche
nn unterschiede

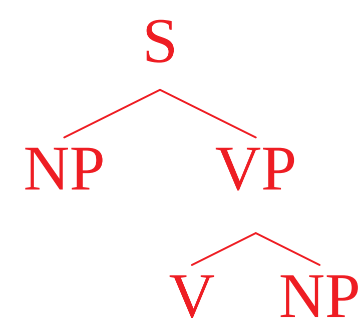
Paraphrase: [pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]</td><td>STEM: be

SPINE: 

VOICE: active
SUBJECT: there
OBJECT: [3]
WH: NULL
MODALS: NULL
INFL: are
MOD1: post-verb
MOD2:
MOD3:</td></tr>
</table>

<table>
<tr><th>German Clause</th><th>English AEP</th></tr>
<tr><td>
s pp-mo[1] appr zwischen
piat beiden
nn gesetzen
vvfin-hd bestehen
adv-mo[2] also
np-sb[3] adja erhebliche
adja rechtliche
$, ,
adja praktische
kon und
adja wirtschaftliche
nn unterschiede

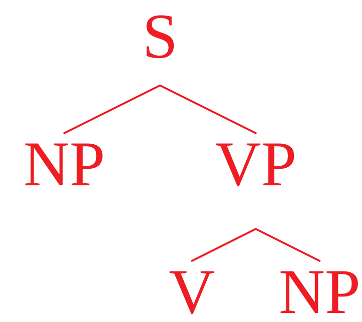
Paraphrase: [pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]
</td><td>
STEM: be

SPINE:


VOICE: active
SUBJECT: there
OBJECT: [3]
WH: NULL
MODALS: NULL
INFL: are
MOD1: post-verb
MOD2: pre-subject
MOD3:
</td></tr>
</table>

<table>
<tr><th>German Clause</th><th>English AEP</th></tr>
<tr><td>
s pp-mo[1] appr zwischen
piat beiden
nn gesetzen
vvfin-hd bestehen
adv-mo[2] also
np-sb[3] adja erhebliche
adja rechtliche
$, ,
adja praktische
kon und
adja wirtschaftliche
nn unterschiede

Paraphrase: [pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]
</td><td>
STEM: be

SPINE:


VOICE: active
SUBJECT: there
OBJECT: [3]
WH: NULL
MODALS: NULL
INFL: are
MOD1: post-verb
MOD2: pre-subject
MOD3: null
</td></tr>
</table>

Linear Models

- x is a German clause, $y = \langle d_1, \ldots, d_n \rangle$ is an English AEP. (d_j is the j'th decision.)

- Define $y^* = \arg\max_y Score(x, y)$ where

$$Score(x, y) = \sum_{j=1}^{n} \bar{\phi}(x, d_1 \ldots d_{j-1}, d_j) \cdot \bar{\theta}$$

Linear Models

- x is a German clause, $y = \langle d_1, \ldots, d_n \rangle$ is an English AEP.

- Define $y^* = \arg\max_y Score(x, y)$ where

$$Score(x, y) = \sum_{j=1}^{n} \bar{\phi}(x, d_1 \ldots d_{j-1}, d_j) \cdot \bar{\theta}$$

- Search for the $\arg\max$ using **beam search**
  (Closely related to history-based models of Jelinek et al., 1994, Ratnaparkhi 1997)

- Train the parameters $\bar{\theta}$ using the **perceptron algorithm**
  (Similar to Collins and Roark, 2004)

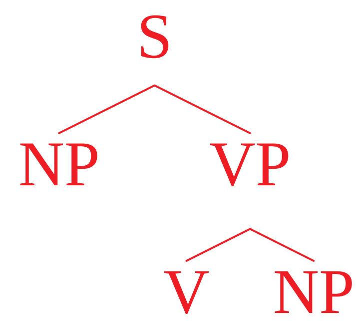
Example Features: Prediction of the Stem

Features consider the verb stem, plus:

- Main verb in the German clause
- Each of the verbs in the German clause
- The entire sequence of German verbs
- The “spine” of the German clause
- The entire parse tree for the German clause
- The rank of the English stem paired with the German main verb, in a lexicon compiled by GIZA++

<table>
<tr><th>German Clause</th><th>English AEP</th></tr>
<tr><td>
s pp-mo[1] appr zwischen
piat beiden
nn gesetzen
vvfin-hd bestehen
adv-mo[2] also
np-sb[3] adja erhebliche
adja rechtliche
$, ,
adja praktische
kon und
adja wirtschaftliche
nn unterschiede

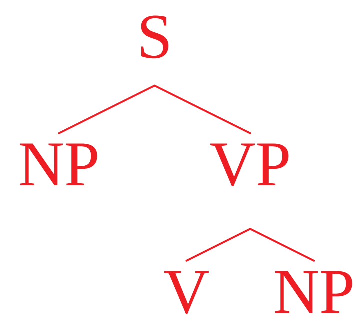
Paraphrase: [pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]
</td><td>
STEM: be

SPINE:


VOICE: active
SUBJECT: there
OBJECT:
WH:
MODALS:
INFL:
MOD1:
MOD2:
MOD3:
</td></tr>
</table>

<table>
<tr><th>German Clause</th><th>English AEP</th></tr>
<tr><td>s pp-mo[1] appr zwischen
piat beiden
nn gesetzen
vvfin-hd bestehen
adv-mo[2] also
np-sb[3] adja erhebliche
adja rechtliche
$, ,
adja praktische
kon und
adja wirtschaftliche
nn unterschiede

Paraphrase: [pp-mo between the two pieces of legislation] exist so [np-sb significant legal, practical and economic differences]</td>
<td>STEM: be

SPINE: 

VOICE: active
SUBJECT: there
OBJECT: [3]
WH:
MODALS:
INFL:
MOD1:
MOD2:
MOD3:</td></tr>
</table>

Outline

- An architecture for tree-to-tree translation
- Training the model using the perceptron algorithm
- Experiments
- Conclusions

Experiments

- Training data: 750,000 English/German sentences from Europarl
- Parsed English side using parser from (Collins, 1999)
- Parsed German side using parser from (Dubey, 2005)
- Extracted 450,000 AEP examples, trained a perceptron model
- In addition, trained Pharaoh (Koehn, Och and Knight, 2003) on the data: used this for generating n-best list for modifier translations

Results

- 2000 sentences from Europarl, compared the AEP system against Pharaoah
- Baseline: translated all sentences using Pharaoh
- AEP: Translated 1335 sentences using AEP system, remainder using Pharaoah (did not translate sentences which could not be broken down into a sequence of German clauses that spanned the sentence)
- Results: Pharaoh = 25.26 BLEU, AEP = 23.96 BLEU
- Human evaluation of 100 translations where the two systems differed

	Equal quality	Pharaoh Better	AEP better
Annotator 1	62	22	16
Annotator 2	37	32	31

Recent Work: Finite State Machines

- Current approach generates each modifier translation independently of AEP, and of other modifiers. In particular, lacks a *language model* across modifier boundaries.

- Finite state approach: for each modifier, create a *lattice* of possible translations. Combine this with an AEP to create a *lattice* of possible clause translations. Can then compose an n-gram language model with the lattice.

- Can also make use of n-best AEPs with this method

- Preliminary results: optimize combination of AEP (perceptron) score, language model, Pharaoah scores.

	BLEU
First best	20.26
Finite State	22.03
Phrase-based	23.11

(**First best** system takes 1-best AEP, 1-best modifiers when forming a translation.)

Future Work: A language model for AEPs

- Can induce a prior distribution over AEP structures from a large English corpus

$$P(d_1, d_2, \ldots d_n) = \prod_{j=1}^{n} P(d_j \mid d_1 \ldots d_{j-1})$$

- Incorporate each term

$$\log P(d_j \mid d_1 \ldots d_{j-1})$$

as a feature in the j'th decision for the perceptron

Future Work: Improved Extraction of AEPs from Training Data

- Current method for AEP extraction is fairly simple, but is a heuristic method
- Improved method might make use of EM, or other algorithm for inducing alignments between/within clauses
- A related point: extend the AEP model to predict larger pieces of structure (e.g., an EP plus 1 or more modifiers)

Conclusions

- Main contributions:
  - A translation architecture based on aligned extended projections
  - A perceptron-based learning method for AEPs
- Motivation:
  - AEPs encapsulate complex correspondances between German/English at the clausal level
  - Learning method is highly flexible in terms of the features that can be included in the model