TAG-based Structured Prediction Models for Parsing and Machine Translation

Michael Collins

MIT

Joint work with Xavier Carreras, Terry Koo
Syntactic Structures

S
  /\ 
 NP  VP
   /\   /
  N   NP  PP
    /\  /\ 
   V  NP  N
  /\    /\ 
 John saw Mary in Helsinki
Canadian Utilities had 1988 revenue of C$ 1.16 billion, mainly from its natural gas and electric utility businesses in Alberta, where the company serves about 800,000 customers.
Statistical Machine Translation

► **Data:** a bilingual parallel corpus

<table>
<thead>
<tr>
<th>German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wiederaufnahme der Sitzungsperiode.</td>
<td>Resumption of the session.</td>
</tr>
<tr>
<td>Gibt es Einwände?</td>
<td>Are there any comments?</td>
</tr>
<tr>
<td>Wissenschaftlich betrachtet haben Sie recht.</td>
<td>Scientifically you are right.</td>
</tr>
<tr>
<td>Sie sind äußerst wichtig.</td>
<td>They are extremely important.</td>
</tr>
<tr>
<td>Das Wort hat Herr Simpson.</td>
<td>Mr Simpson has the floor.</td>
</tr>
<tr>
<td>Bedauerlicherweise wurde dies nicht eingehalten.</td>
<td>Sadly, that has not been the case.</td>
</tr>
<tr>
<td>Vielen Dank, Herr Simpson.</td>
<td>Thank you very much, Mr Simpson.</td>
</tr>
</tbody>
</table>

► **Goal:** learn a model that can predict an English translation given a German sentence
In a strategic assessment of the Afghanistan conflict given to Robert Gates August 30, revealed by The Washington Post Monday, General McChrystal warned that without an increase in military resources in Afghanistan, the coalition might undergo a “failure.”
Problem 2: Translation as Parsing

wir müssen diese kritik ernst nehmen
(we must these criticisms seriously take)
The choice of grammar formalism implies a decomposition of parse trees into smaller units.

This choice is critical to:
1. Representations that can be used
2. Computational efficiency of underlying algorithms

This talk: a Tree Adjoining Grammar (TAG) formalism for parsing and translation
Outline

- A Tree Adjoining Grammar (TAG) formalism
- A TAG-based discriminative parser
- A TAG-based translation model
In Tree Adjoining Grammar (TAG, Joshi, 1985) the grammar is defined by a set of elementary trees.

Our elementary trees are Spines (See also Shen and Joshi, 2005):
A Combination Operation: *Sister Adjunction*

*Sister adjunctions* are used to combine spines to form trees.

![Diagram of sister adjunction in a tree structure]

An adjunction operation attaches:

- A **modifier** spine
- To some **position** of a **head** spine
A Combination Operation: *Sister Adjunction*

*Sister adjunctions* are used to combine spines to form trees.

An adjunction operation attaches:
- A *modifier* spine
- To some *position* of a *head* spine
A Combination Operation: *Sister Adjunction*

*Sister adjunctions* are used to combine spines to form trees.

An adjunction operation attaches:
- A *modifier* spine
- To some *position* of a *head* spine
The Decomposition into Spines and Adjunctions

[Diagram showing the syntax tree for the sentence: Mary eats the cake with almonds.

Tree on the left:
- S
  - NP
    - n
      - Mary
  - VP
    - v
      - eats
    - NP
      - d
        - the
      - n
        - cake
    - PP
      - p
        - with
      - NP
        - n
          - almonds

Tree on the right:
- S
  - NP
    - n
      - Mary
  - VP
    - v
      - eats
    - NP
      - d
        - the
      - n
        - cake
    - PP
      - p
        - with
      - NP
        - n
          - almonds]
A Little More Formally....

Each spine has a separate left/right weighted finite-state automaton (HMM) at each level of the tree (in this case $S$, $VP$)

The automata generate sequences of modifier spines at each level of the tree

Parsing complexity: $O(n^3 G)$ where $n$ is the length of the string, $G$ is a grammar constant (Eisner 2000)
Advantages of TAG

- Lexical entries naturally capture constraints associated with lexical items

```
S
| VP
| V
| saw
```

- Probabilities/costs can be associated with combination operations:

```
S + NP ⇒ S
| VP
| N
| cake
| V
| eats

S
| VP
| V
| cake
| eats
| N
| cake
```
Outline

- A Tree Adjoining Grammar (TAG) formalism
- A TAG-based discriminative parser
- A TAG-based translation model
Structured Prediction Models for Parsing

- Conditional random fields (CRFs), and other discriminative models, are a powerful alternative to HMMs
  - A key strength: flexible representations

- Can we generalize CRF-style models to parsing?
Features on Adjunctions

Feature vectors $f(x, h, m, \sigma_h, \sigma_m, \text{POS})$ where
- $x$ is the sentence
- $h = 3$ (index of head word), $m = 5$ (index of modifier word)
- $\sigma_h$ and $\sigma_m$ are the head and modifier spines
- POS is the position being adjoined into (e.g., VP)
Features on Adjunctions

Feature vectors $f(x, h, m, \sigma_h, \sigma_m, \text{POS})$ where

- $x$ is the sentence
- $h = 3$ (index of head word), $m = 5$ (index of modifier word)
- $\sigma_h$ and $\sigma_m$ are the head and modifier spines
- POS is the position being adjoined into (e.g., VP)
Features on Adjunctions

Feature vectors $f(x, h, m, \sigma_h, \sigma_m, \text{POS})$ where

- $x$ is the sentence
- $h = 3$ (index of head word), $m = 5$ (index of modifier word)
- $\sigma_h$ and $\sigma_m$ are the head and modifier spines
- POS is the position being adjoined into (e.g., VP)
Features on Adjunctions

Feature vectors $f(x, h, m, \sigma_h, \sigma_m, \text{POS})$ where

- $x$ is the sentence
- $h = 3$ (index of head word), $m = 5$ (index of modifier word)
- $\sigma_h$ and $\sigma_m$ are the head and modifier spines
- POS is the position being adjoined into (e.g., VP)
Features on Adjunctions

Feature vectors \( f(x, h, m, \sigma_h, \sigma_m, \text{POS}) \) where

- \( x \) is the sentence
- \( h = 3 \) (index of head word), \( m = 5 \) (index of modifier word)
- \( \sigma_h \) and \( \sigma_m \) are the head and modifier spines
- \( \text{POS} \) is the position being adjoined into (e.g., VP)
Features on Adjunctions

Feature vectors $f(x, h, m, \sigma_h, \sigma_m, \text{POS})$ where

- $x$ is the sentence
- $h = 3$ (index of head word), $m = 5$ (index of modifier word)
- $\sigma_h$ and $\sigma_m$ are the head and modifier spines
- POS is the position being adjoined into (e.g., VP)
Features on Adjunctions

Feature vectors \( f(x, h, m, \sigma_h, \sigma_m, \text{POS}) \) where

- \( x \) is the sentence
- \( h = 3 \) (index of head word), \( m = 5 \) (index of modifier word)
- \( \sigma_h \) and \( \sigma_m \) are the head and modifier spines
- \( \text{POS} \) is the position being adjoined into (e.g., VP)
A TAG-Based Model

- Goal: map an input sentence $x$ to a parse tree $y$

- Model form:

$$y^* = \arg \max_y \sum_{r \in y} w \cdot f(x, r)$$

where each $r$ is a tuple $\langle h, m, \sigma_h, \sigma_m, \text{POS} \rangle$ representing a combination of two spines in $y$

- Parameter estimation: we used the averaged perceptron

- The inference problem: How to compute $y^*$?
Inference: Key Points

- Dynamic programming algorithms can be applied to the TAG grammars

- Exact inference is still very expensive

- A solution: coarse-to-fine dynamic programming (e.g., (Charniak, 1997; Charniak and Johnson, 2005))
  - Use a first-pass, simple, computationally-cheap model to restrict the search space of the full model
Dependency Structures

- Directed arcs represent *dependencies* between a *head word* and a *modifier word*.

- Dependency parsing models of McDonald et al. (2005, 2006):

\[
y^* = \arg \max_y \sum_{r \in y} w \cdot f(x, r)
\]

where each \( r \) is a tuple \( \langle h, m \rangle \) representing a dependency from modifier \( m \) to head \( h \).
A dependency structure augmented with spines, and attachment positions
Coarse-to-fine Dynamic Programming

Coarse-to-fine approach: we only allow the full TAG model to consider dependencies that have high probability under a (simple) dependency model

The simple model estimates dependency probabilities in $O(n^3 G)$ time, where $G \approx 60$ is the number of non-terminals (i.e., VP, NP, S, etc.)
Test results on WSJ data

<table>
<thead>
<tr>
<th>FULL PARSERS</th>
<th>precision</th>
<th>recall</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Petrov et al. 2007</td>
<td></td>
<td></td>
<td>88.3</td>
</tr>
<tr>
<td>Finkel et al. 2008</td>
<td>88.2</td>
<td>87.8</td>
<td>88.0</td>
</tr>
<tr>
<td>Charniak 2000</td>
<td>89.5</td>
<td>89.6</td>
<td>89.6</td>
</tr>
<tr>
<td>Petrov &amp; Klein 2007</td>
<td>90.2</td>
<td>89.9</td>
<td>90.1</td>
</tr>
<tr>
<td><strong>this work</strong></td>
<td><strong>91.4</strong></td>
<td><strong>90.7</strong></td>
<td><strong>91.1</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RERANKERS</th>
<th>precision</th>
<th>recall</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collins 2000</td>
<td>89.9</td>
<td>89.6</td>
<td>89.8</td>
</tr>
<tr>
<td>Charniak &amp; Johnson 2005</td>
<td></td>
<td></td>
<td><strong>91.4</strong></td>
</tr>
</tbody>
</table>
## Effect of the Beam (Validation Data)

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>1st stage active</th>
<th>1st stage cov.</th>
<th>2nd stage orac.</th>
<th>2nd stage speed</th>
<th>$F_1$</th>
<th>error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-4}$</td>
<td>0.07</td>
<td>97.7</td>
<td>97.0</td>
<td>5:15</td>
<td>8.9</td>
<td></td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>0.16</td>
<td>98.5</td>
<td>97.9</td>
<td>11:45</td>
<td>8.4</td>
<td></td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>0.34</td>
<td>99.0</td>
<td>98.5</td>
<td>21:50</td>
<td>8.0</td>
<td></td>
</tr>
</tbody>
</table>

We can discard $99.6\%$ of the possible adjunctions and retain $98.5\%$ of the correct syntactic constituents.
Outline

- A Tree Adjoining Grammar (TAG) formalism
- A TAG-based discriminative parser
- A TAG-based translation model
Phrase-based Systems: Derivations

In wenigen Tagen finden Parlamentswahlen in Slowenien statt

Translation involves:

1. Segmenting the input into phrases, and choosing a translation for each phrase
2. Choosing an ordering of the resulting English phrases
Phrase-based Systems: Derivations

Translation involves:

1. Segmenting the input into phrases, and choosing a translation for each phrase
2. Choosing an ordering of the resulting English phrases
Phrase-based Systems: Derivations

[In wenigen] [Tagen] [finden] [Parlamentswahlen] [in Slowenien] [statt]
[In a few] [days] [take] [elections] [in Slovenia] [place]

Translation involves:

1. Segmenting the input into phrases, and choosing a translation for each phrase
2. Choosing an ordering of the resulting English phrases
Phrase-base Systems: a Phrase Table

auch ⇒ also
auf nationaler und ⇒ at national and
bereits ⇒ already
dass ⇒ that
der kommission ⇒ the commission
des besitzstandes ⇒ of the acquis
die wichtigste ⇒ the most important
gemeinschaftspolitiken ⇒ community policies
im dezember in nizza ⇒ in december in nice
in diesem bericht enthaltenen ⇒ contained in this report
ist notwendig und ⇒ is necessary and
menschenrechte ⇒ human rights
oppositionsparteien und ⇒ opposition parties and
positiven auswirkungen der ⇒ positive effects of
trennlinie ⇒ dividing line
umsetzung der menschenrechte ⇒ implementation of human rights
und die ⇒ and the
wird schrittweise ⇒ should be gradually
zu beachten haben ⇒ to bear in mind
<table>
<thead>
<tr>
<th>German Word or Phrase</th>
<th>English Translation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>auch</td>
<td>also</td>
<td>0.73</td>
</tr>
<tr>
<td>auf nationaler und</td>
<td>at national and</td>
<td>0.34</td>
</tr>
<tr>
<td>bereits</td>
<td>already</td>
<td>0.65</td>
</tr>
<tr>
<td>dass</td>
<td>that</td>
<td>0.92</td>
</tr>
<tr>
<td>der kommission</td>
<td>the commission</td>
<td>0.85</td>
</tr>
<tr>
<td>des besitzstandes</td>
<td>of the acquis</td>
<td>0.56</td>
</tr>
<tr>
<td>die wichtigste</td>
<td>the most important</td>
<td>0.44</td>
</tr>
<tr>
<td>gemeinschaftspolitiken</td>
<td>community policies</td>
<td>0.31</td>
</tr>
<tr>
<td>im dezember in nizza</td>
<td>in december in nice</td>
<td>0.37</td>
</tr>
<tr>
<td>in diesem bericht enthaltenen</td>
<td>contained in this report</td>
<td>0.81</td>
</tr>
<tr>
<td>ist notwendig und</td>
<td>is necessary and</td>
<td>0.94</td>
</tr>
<tr>
<td>menschenrechte</td>
<td>human rights</td>
<td>0.78</td>
</tr>
<tr>
<td>oppositionsparteien und</td>
<td>opposition parties and</td>
<td>0.53</td>
</tr>
<tr>
<td>positiven auswirkungen der</td>
<td>positive effects of</td>
<td>0.58</td>
</tr>
<tr>
<td>trennlinie</td>
<td>dividing line</td>
<td>0.67</td>
</tr>
<tr>
<td>umsetzung der menschenrechte</td>
<td>implementation of human rights</td>
<td>0.96</td>
</tr>
<tr>
<td>und die</td>
<td>and the</td>
<td>0.89</td>
</tr>
<tr>
<td>wird schrittweise</td>
<td>should be gradually</td>
<td>0.85</td>
</tr>
<tr>
<td>zu beachten haben</td>
<td>to bear in mind</td>
<td>0.44</td>
</tr>
</tbody>
</table>
on all these subjects, the brok report confines itself to discussing adaptation and reform.
bei all diesen problemen beschränkt sich der bericht brok darauf, von anpassung oder reformen zu sprechen.

Paraphrase: on all these subjects confines itself the report brok on adaptation and reform to speak

on all these subjects, the brok report confines itself to discussing adaptation and reform.
Word-order Differences

bei all diesen problemen beschränkt sich der bericht brok darauf, von anpassung oder reformen zu sprechen.

Paraphrase: on all these subjects confines itself the report brok on adaptation and reform to speak

on all these subjects, the brok report confines itself to discussing adaptation and reform.

Translation: with all these problems is limited to the report brok to talk about reform or adjustment.
Phrase-based Translation with TAG operations

There is no higher level discrimination between hierarchy of discrimination

es gibt keine hierarchie der diskriminierung

segmentation + s-phrase selection + adjunctions
Phrase-based Translation with TAG operations

There is no higher level of discrimination between a higher level of discrimination. There is no higher level of discrimination between a higher level of discrimination. There is no higher level of discrimination between a higher level of discrimination. There is no higher level of discrimination between a higher level of discrimination.

es gibt keine hierarchie der diskriminierung

segmentation + s-phrase selection + adjunctions
Reordering via Non-Projective Operations

we must also take these criticisms seriously
diese kritik ernst nehmen

wir müssen auch

segmentation + s-phrase selection + non-projective adjunctions
Reordering via Non-Projective Operations

we must also take these criticisms seriously take

wir müssen auch diese kritik ernst nehmen
take these criticisms

segmentation + s-phrase selection + non-projective adjunctions
Reordering via Non-Projective Operations

segmentation + s-phrase selection + non-projective adjuncions
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

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
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

Training example = source sentence + English sentence + English parse tree

We use phrasal entries from a standard phrase-based approach
A derivation:

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

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

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

Probabilities track:

1. Phrase probabilities
   (how likely is *diese kritik* to be translated as *these criticisms*?)
2. Grammatical relations
   (how likely is *take* to take *criticisms* as a direct object?)
3. Relationship to German string (how likely is ⟨*kritik, nehmen*⟩ to result in a verb-object relation in English?)
4. Surface trigram probabilities: e.g., $P(\text{criticisms} \mid \text{take, these})$
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:**

```
S → NP → VP → ADVP
we must also
dauswählen auch
diese kritik ernst nehmen
```

**Violated:**

```
S → NP → VP → DT → NP → PP → NP
there is no hierarchy of discrimination
es gibt keine hierarchie der diskriminierung
```
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
Beam Search Decoding

0. Data structures: \( Q_i \) for \( i = 1 \ldots n \) is a set of hypotheses for each length \( i \), \( S \) is a set of chart entries

1. \( S \leftarrow \emptyset \)

2. Initialize \( Q_1 \ldots Q_n \) with basic chart entries derived from phrase entries

3. For \( i = 1 \ldots n \)

4. For any \( A \in \text{BEAM}(Q_i) \)

5. If \( S \) contains a chart entry with the same signature as \( A \), and which has a higher inside score,

6. continue

7. Else

8. Add \( A \) to \( S \)

9. For any chart entry \( C \) that can be derived from \( A \) together with another chart entry \( B \in S \), add \( C \) to the set \( Q_j \) where \( j = \text{length}(C) \)

10. Return \( Q_n \), a set of items of length \( n \)
The Definition of BEAM

(BEAM) Given $Q_i$, define $Q_{i,j}$ for $j = 1 \ldots n$ to be the subset of items in $Q_i$ which have their $j$’th bit equal to one (i.e., have the $j$’th source language word translated). Define $Q'_{i,j}$ to be the $N$ highest scoring elements in $Q_{i,j}$. Then $\text{BEAM}(Q_i) = \bigcup_{j=1}^{n} Q'_{i,j}$. 

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

Development:

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax-based</td>
<td>25.2</td>
</tr>
<tr>
<td>Syntax (no disc. model)</td>
<td>23.7 (-1.5)</td>
</tr>
<tr>
<td>Syntax (no $\pi$-c constraint)</td>
<td>24.4 (-0.8)</td>
</tr>
</tbody>
</table>

Test:

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrase-based system (Pharaoh)</td>
<td>24.58</td>
</tr>
<tr>
<td>Syntax-based system</td>
<td>25.04 (+0.46)</td>
</tr>
</tbody>
</table>

significant ($p = 0.021$) under paired bootstrap resampling [Koehn 04]
close to significant ($p = 0.058$) under the sign test [Collins et al. 05]
Experiments
German to English using Europarl data (750K training sentences)

Development:

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax-based</td>
<td>25.2</td>
</tr>
<tr>
<td>Syntax (no disc. model)</td>
<td>23.7 (-1.5)</td>
</tr>
<tr>
<td>Syntax (no $\pi$-c constraint)</td>
<td>24.4 (-0.8)</td>
</tr>
</tbody>
</table>

Test:

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrase-based system (Pharaoh)</td>
<td>24.58</td>
</tr>
<tr>
<td>Syntax-based system</td>
<td>25.04 (+0.46)</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>Syntax</th>
<th>PB</th>
<th>=</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax</td>
<td>51</td>
<td>3</td>
<td>7</td>
<td>61</td>
</tr>
<tr>
<td>PB</td>
<td>1</td>
<td>25</td>
<td>11</td>
<td>37</td>
</tr>
<tr>
<td>=</td>
<td>21</td>
<td>14</td>
<td>67</td>
<td>102</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>42</td>
<td>85</td>
<td>200</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Syntax</th>
<th>PB</th>
<th>=</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax</td>
<td>51</td>
<td>3</td>
<td>7</td>
<td>61</td>
</tr>
<tr>
<td>PB</td>
<td>1</td>
<td>25</td>
<td>11</td>
<td>37</td>
</tr>
<tr>
<td>=</td>
<td>21</td>
<td>14</td>
<td>67</td>
<td>102</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>42</td>
<td>85</td>
<td>200</td>
</tr>
</tbody>
</table>

both results are significant with $p < 0.05$ under the sign test
Reference: now, however, we are seeing that president putin is pursuing a policy of openness towards the west.

Phrase-based: we are, however, now that president putin a policy of openness to the west out of blackmail

Syntax: now, however, we see that mr president putin is pursuing a policy of openness towards the west.
Reference: on all these subjects, the brok report confines itself to discussing adaptation and reform.

Phrase-based: in all these issues is limited to the brok report, adjustment or reforms to speak.

Syntax: the brok report is limited to speak of adjustment or reforms in all these issues.
Reference: I believe that deferring the issue would be the worst possible option, both for the citizens of Europe and for the citizens of the candidate countries.

Phrase-based: I believe, however, that postpone a decision would be the worst possible both for the citizens of Europe, as well as for the citizens of the candidate countries.

Syntax: I believe, however, that a postponement would be the worst possible choice both for the citizens of the union and for the citizens of the candidate countries.
Syntactified BLEU Scores

► Regular BLEU scores:

SYNTAX = 24.96, 59.6/31.5/19.0/11.8 (BP=0.980)
PHRASEB = 24.48, 59.7/30.7/18.3/11.3 (BP=0.986)

► Adding syntactic annotations:

parliament: S+VP+NP agrees: ROOT+ROOT+S

► Syntactified BLEU scores:

SYNTAX = 18.83, 46.5/23.7/14.1/8.8 (BP=0.980)
PHRASEB = 16.93, 43.4/21.4/12.4/7.5 (BP=0.986)
Future Work

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
Future Work

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
Summary

- A TAG-based formalism. Key points:
  - Combines dependency and constituency based representations
  - Allows relatively efficient parsing algorithms

- A TAG-based discriminative parser. Key points: feature-vector representations of TAG adjunctions, coarse-to-fine inference

- A TAG-based translation model. Key points: non-projective parsing operations, a discriminative dependency model