Rule Extraction for Machine Translation

Andreas Maletti

Institute of Computer Science
Universität Leipzig, Germany

on leave from: Institute for Natural Language Processing
Universität Stuttgart, Germany

maletti@ims.uni-stuttgart.de

Darmstadt — September 10, 2014
Main notions

<table>
<thead>
<tr>
<th>Machine translation (MT)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Automatic</strong> natural language translation (by a computer)</td>
</tr>
<tr>
<td>as opposed to:</td>
</tr>
<tr>
<td>- manual translation</td>
</tr>
<tr>
<td>- computer-aided translation (e.g., translation memory)</td>
</tr>
</tbody>
</table>
## Main notions

### Machine translation (MT)

*Automatic* natural language *translation* (by a computer)

As opposed to:
- manual translation
- computer-aided translation (e.g., translation memory)

### Statistical machine translation (SMT)

MT using systems *automatically* obtained from *translations*

As opposed to:
- rule-based machine translation (old) SYSTRAN
- example-based machine translation translation by analogy
## Short history

### Timeline

<table>
<thead>
<tr>
<th></th>
<th>Dark age (60s–90s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rule-based systems (e.g., SYSTRAN)</td>
</tr>
<tr>
<td></td>
<td>CHOMSKYAN approach</td>
</tr>
<tr>
<td></td>
<td>perfect translation, poor coverage</td>
</tr>
</tbody>
</table>
Short history

Timeline

1. Dark age (60s–90s)
   - rule-based systems (e.g., SYSTRAN)
   - CHOMSKYAN approach
   - perfect translation, poor coverage

2. Reformation (1991–present)
   - phrase-based and syntax-based systems
   - statistical approach
   - cheap, automatically trained
# Short history

## Timeline

<table>
<thead>
<tr>
<th></th>
<th>Dark age (60s–90s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rule-based systems (e.g., SYSTRAN)</td>
</tr>
<tr>
<td></td>
<td>CHOMSKYAN approach</td>
</tr>
<tr>
<td></td>
<td>perfect translation, poor coverage</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Reformation (1991–present)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>phrase-based and syntax-based systems</td>
</tr>
<tr>
<td></td>
<td>statistical approach</td>
</tr>
<tr>
<td></td>
<td>cheap, automatically trained</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Potential future</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>semantics-based systems (e.g., FRAMESNET-based)</td>
</tr>
<tr>
<td></td>
<td>semi-supervised, statistical approach</td>
</tr>
<tr>
<td></td>
<td>basic understanding of translated text</td>
</tr>
</tbody>
</table>
Examples

Applications
- Technical manuals

Example (An mp3 player)

The synchronous manifestation of lyrics is a procedure for can broadcasting the music, waiting the mp3 file at the same time showing the lyrics.
Examples

Applications

- Technical manuals

Example (An mp3 player)

With this kind of method, the equipments that support synchronous function of broadcast make use of document create setup. You can pass the document contents that broadcast.

That procedure returns offerings to modify, delete, and stick, etc. edit function.
Examples

Applications

- Technical manuals

Example (An mp3 player)

That procedure returns offerings to have to modify, and delete, and stick top, keep etc. edit function.
Example (Hotel Uppsala, Sweden)


— We stayed in rooms classified as “superior” because they had been renovated in the last year or two. Our rooms had wood floors and were roomy. You didn’t have to walk sideways to move around.
Examples

Applications

- Technical manuals
- TripAdvisor
- US military

Example (JONES, SHEN, HERZOG 2009)

Soldier: Okay, what is your name?
Local: Abdul.
Soldier: And your last name?
Local: Al Farran.
Examples

Applications

- Technical manuals
- US military

Example (JONES, SHEN, HERZOG 2009)

Soldier: Okay, what is your name?
Local: Abdul.
Soldier: And your last name?
Local: Al Farran.

Speech-to-text machine translation

Soldier: Okay, what’s your name?
Local: milk a mechanic and I am here I mean yes
Examples

Applications
- Technical manuals
- TripAdvisor
- US military

Example (JONES, SHEN, HERZOG 2009)

| Soldier:  | Okay, what is your name? |
| Local:    | Abdul.                   |
| Soldier:  | And your last name?     |
| Local:    | Al Farran.              |

Speech-to-text machine translation

| Soldier:  | Okay, what’s your name? |
| Local:    | milk a mechanic and I am here I mean yes |
| Soldier:  | What is your last name? |
| Local:    | every two weeks my son’s name is ismail |
Examples

Applications

- Technical manuals
- tripadvisor
- US military
- MSDN, Knowledge Base
- ...
Standard pipeline

Input $\rightarrow$ Translation model $\rightarrow$ Language model $\rightarrow$ Output

(the models are often integrated in practice)
Standard pipeline

Schema

Input $\rightarrow$ Translation model $\rightarrow$ Language model $\rightarrow$ Output

(the models are often integrated in practice)

Required resources

- bilingual text (sentences in both languages) 1.5M sent.
Standard pipeline

Schema

Input $\rightarrow$ Translation model $\rightarrow$ Language model $\rightarrow$ Output

(the models are often integrated in practice)

Required resources

- bilingual text (sentences in both languages) 1.5M sent.
- monolingual text (in target language) 44M sent.
Word Alignment

**English-German example**

We can help countries catch up, but not by putting their neighbours on hold.

Wir können Ländern beim Aufholen helfen, aber nicht indem wir ihre Nachbarn in den Wartesaal schicken.

**English-Russian example**

I was the one who sat down and copied them.

Я был единственным, кто занялся копированием их демонстрационную кассету.
Parallel Corpus

**EUROPARL German-English parallel corpus**

- 1,920,209 parallel sentences
- 44,548,491 words in German
- 47,818,827 words in English
- sentence-aligned, but not word-aligned
- from parliament proceedings
Könnten Sie mir eine Auskunft zu Artikel 143 im Zusammenhang mit der Unzulässigkeit geben?

I would like your advice about Rule 143 concerning inadmissibility.

Algorithm

1. phrase pair \([j, j'], [i, i']\) consistently aligned if
   - \(\ell' \in [i, i']\) for all \(\ell \in [j, j']\) and \((\ell, \ell') \in A\)
   - \(\ell \in [j, j']\) for all \(\ell' \in [i, i']\) and \((\ell, \ell') \in A\)
Könnten Sie mir eine Auskunft zu Artikel 143 im Zusammenhang mit der Unzulässigkeit geben?

I would like your advice about Rule 143 concerning inadmissibility.

Algorithm

1. Phrase pair ([j, j'], [i, i']) consistently aligned if
   - ℓ' ∈ [i, i'] for all ℓ ∈ [j, j'] and (ℓ, ℓ') ∈ A
   - ℓ ∈ [j, j'] for all ℓ' ∈ [i, i'] and (ℓ, ℓ') ∈ A

2. Extract all consistently aligned phrase pairs

3. (restrict length of phrases based on corpus size)
Phrase-based Models

Formally:

\[
([1,1], [2,3]) \quad ([2,2], [4,4]) \quad ([3,3], [1,1])
\]

\[
([4,5], [5,5]) \quad ([6,6], [6,6]) \quad ([7,7], [7,7])
\]

\[
([8,8], [8,8]) \quad ([9,11], [9,9]) \quad ([12,13], [10,10])
\]
Phrase-based Models

Formally:

\[
([1,1], [2,3]) \quad ([2,2], [4,4]) \quad ([3,3], [1,1])
\]
\[
([4,5], [5,5]) \quad ([6,6], [6,6]) \quad ([7,7], [7,7])
\]
\[
([8,8], [8,8]) \quad ([9,11], [9,9]) \quad ([12,13], [10,10])
\]
Phrase-based Models

Formally:

\[
\begin{align*}
([1,1], [2,3]) & \quad ([2,2], [4,4]) & \quad ([3,3], [1,1]) \\
([4,5], [5,5]) & \quad ([6,6], [6,6]) & \quad ([7,7], [7,7]) \\
([8,8], [8,8]) & \quad ([9,11], [9,9]) & \quad ([12,13], [10,10])
\end{align*}
\]
Phrase-based Models

Formally:

\[
\begin{align*}
&([1,1], [2,3]) \quad ([2,2], [4,4]) \quad ([3,3], [1,1]) \\
&([4,5], [5,5]) \quad ([6,6], [6,6]) \quad ([7,7], [7,7]) \\
&([8,8], [8,8]) \quad ([9,11], [9,9]) \quad ([12,13], [10,10])
\end{align*}
\]
Phrase-based Models

Formally:

$$([1,1], [2,3]) \quad ([2,2], [4,4]) \quad ([3,3], [1,1])$$

$$([4,5], [5,5]) \quad ([6,6], [6,6]) \quad ([7,7], [7,7])$$

$$([8,8], [8,8]) \quad ([9,11], [9,9]) \quad ([12,13], [10,10])$$
Könnten Sie mir eine Auskunft zu Artikel 143 im Zusammenhang mit der Unzulässigkeit geben?

I would like your advice about Rule 143 concerning inadmissibility.

Formally:

([1,1], [2,3])  ([2,2], [4,4])  ([3,3], [1,1])
([4,5], [5,5])  ([6,6], [6,6])  ([7,7], [7,7])
([8,8], [8,8])  ([9,11], [9,9])  ([12,13], [10,10])
Phrase-based Models

Könnten Sie mir eine Auskunft zu Artikel im Zusammenhang mit der Unzulässigkeit geben

I would like your advice about Rule concerning inadmissibility

Formally:

\[
([1,1], [2,3]) \quad ([2,2], [4,4]) \quad ([3,3], [1,1])
\]

\[
([4,5], [5,5]) \quad ([6,6], [6,6]) \quad ([7,7], [7,7])
\]

\[
([8,8], [8,8]) \quad ([9,11], [9,9]) \quad ([12,13], [10,10])
\]
Formally:

\[
([1,1], [2,3]) \quad ([2,2], [4,4]) \quad ([3,3], [1,1])
\]
\[
([4,5], [5,5]) \quad ([6,6], [6,6]) \quad ([7,7], [7,7])
\]
\[
([8,8], [8,8]) \quad ([9,11], [9,9]) \quad ([12,13], [10,10])
\]
Phrase-based Models

Formally:

\[
\begin{align*}
([1,1], [2,3]) & \quad ([2,2], [4,4]) & \quad ([3,3], [1,1]) \\
([4,5], [5,5]) & \quad ([6,6], [6,6]) & \quad ([7,7], [7,7]) \\
([8,8], [8,8]) & \quad ([9,11], [9,9]) & \quad ([12,13], [10,10])
\end{align*}
\]
Phrase-based Models

Formally:

$$[[1,1], [2,3]] \quad [[2,2], [4,4]] \quad [[3,3], [1,1]]$$
$$[[4,5], [5,5]] \quad [[6,6], [6,6]] \quad [[7,7], [7,7]]$$
$$[[8,8], [8,8]] \quad [[9,11], [9,9]] \quad [[12,13], [10,10]]$$

For better readability:

Könnten — would like
mir — I
eine Auskunft — advice
zu — about
Artikel — Rule
143 — 143
im Zusammenhang mit — concerning
der Unzulässigkeit — inadmissibility
Phrase-based Models

Könnten Sie mir eine Auskunft zu Artikel 143 im Zusammenhang mit der Unzulässigkeit geben?

I would like your advice about Rule 143 concerning inadmissibility.

Notes

- these were only minimal phrase pairs
- extract all (sensible) combinations of these
- e.g., ([1, 1], [2, 3]) and ([2, 2], [4, 4]) yield ([1, 2], [2, 4])
Phrase-based Models

Notes

- these were only \textit{minimal} phrase pairs
- extract all (sensible) combinations of these
- e.g., $([1, 1], [2, 3])$ and $([2, 2], [4, 4])$ yield $([1, 2], [2, 4])$
- unaligned words can be added to neighboring phrases
- e.g., $([12, 13], [10, 10])$ extends to $([12, 14], [10, 10])$
Phrase-based Models

Notes

- these were only minimal phrase pairs
- extract all (sensible) combinations of these
  
  e.g., ([1, 1], [2, 3]) and ([2, 2], [4, 4]) yield ([1, 2], [2, 4])

- unaligned words can be added to neighboring phrases
  
  e.g., ([12, 13], [10, 10]) extends to ([12, 14], [10, 10])

Könnten Sie — would like your der Unzulässigkeit geben — inadmissibility
Phrase-based Models

Könnten Sie mir eine Auskunft zu Artikel 143 im Zusammenhang mit der Unzulässigkeit geben?

I would like your advice about Rule 143 concerning inadmissibility.

Alternative representation (rectangles):
Könnten Sie mir eine Auskunft zu Artikel 143 im Zusammenhang mit der Unzulässigkeit geben?

I would like your advice about Rule 143 concerning inadmissibility.

Alternative representation (rectangles):
Phrase-based Models

Könnten Sie mir eine Auskunft zu Artikel 143 im Zusammenhang mit der Unzulässigkeit geben?

I would like your advice about Rule 143 concerning inadmissibility.

Alternative representation (rectangles):
Könnten Sie mir eine Auskunft zu Artikel 143 im Zusammenhang mit der Unzulässigkeit geben.

I would like your advice about Rule 143 concerning inadmissibility.

Alternative representation (rectangles):
## Rule weights

- simple relative frequencies during extraction
- normally different normalizations as features
Phrase-based Models

Rule weights

- simple relative frequencies during extraction
- normally different normalizations as features

Ideally

- weights should be set to utility of the rule for explaining the training data
- would require reprocessing of training data
- EM or similar algorithms available
  → impractical, EM not used
Syntax-based Machine Translation

Syntax-based systems

Input → Parser → Machine translation system → Language model → Output
Syntax-based Machine Translation

CHARNIAK parser: [CHARNIAK, JOHNSON, 2005]

BitPar parser: [SCHMID, 2006]
Yugoslav President Voislav signed for Serbia.

و تولى التوقيع عن صربيا الرئيس اليوغوسلافي فويسلاف

Translit.: w twlY AltwqyE En SrbyA Alr}ys Alywgws|Afy fwys|Af.

And then the matter was decided, and everything was put in place.

ف كان ان تم الخسم و وضعت الأمور في نصاب ها

Translit.: f kAn An tm AlHsm w wDEt Al>mwr fy nSAb hA.

Below are the male and female winners in the different categories.

و هنا الأوائل و الأوليات في مختلف الفئات

Translit.: w hnA Al>wAjI w Al>wlyAt fy mxlf Alf}At.
Syntax-based Machine Translation

Alignment

Yugoslav → President → Voislav → signed → for → Serbia

w → twlY → AltwqyE → En → SrbyA → Alr}ys → AlywgwsAf → fwysAf
Syntax-based Machine Translation

[Galley et al., 2004]
Syntax-based Machine Translation

- Select next node bottom-up

Rule Extraction for MT

A. Maletti
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

Yugoslav $q_Y$ AlywgwsIAf
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

\[
\begin{align*}
\text{Yugoslav} & \rightarrow q_Y \text{ Alywgws}l\text{Af} \\
\text{President} & \rightarrow q_P \text{ Alr}ys \\
\text{Voislav} & \rightarrow q_v \text{ fwys}l\text{Af} \\
\text{for} & \rightarrow q_f \text{ En} \\
\text{Serbia} & \rightarrow q_s \text{ SrbyA}
\end{align*}
\]
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

\[ \text{Yugoslav}^{q_Y} \text{AlywgwsIAfY} \]
\[ \text{President}^{q_P} \text{Alr} \{ \text{ys} \} \]
\[ \text{Voislav}^{q_V} \text{fwyslAf} \]
\[ \text{for}^{q_f} \text{En} \]
\[ \text{Serbia}^{q_S} \text{SrbyA} \]
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

\[ \text{NML}(q_Y, q_P) \xrightarrow{\text{NML}} \text{NP}(q_P, q_Y) \]
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\text{NML}(q_Y, q_P) \xrightarrow{q_{\text{NML}}} \text{NP}(q_P, q_Y)$$
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

\[
\begin{align*}
\text{NML}(q_Y, q_P) & \rightarrow^{q_{\text{NML}}} \text{NP}(q_P, q_Y) \\
\text{NP}(q_S) & \rightarrow^{q_{\text{NP}}} \text{NP}(q_S) \\
\text{PP}(q_f, q_{\text{NP}}) & \rightarrow^{q_{\text{PP}}} \text{PP}(q_f, q_{\text{NP}}) \\
\text{NP-SBJ}(q_{\text{NML}}, q_V) & \rightarrow^{q_{\text{NP-SBJ}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))
\end{align*}
\]
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

\[
\text{NML}(q_Y, q_P) \xrightarrow{\text{q}_{\text{NML}}} \text{NP}(q_P, q_Y)
\]
\[
\text{NP}(q_S) \xrightarrow{\text{q}_{\text{NP}}} \text{NP}(q_S)
\]
\[
\text{PP}(q_f, q_{\text{NP}}) \xrightarrow{\text{q}_{\text{PP}}} \text{PP}(q_f, q_{\text{NP}})
\]

\[
\text{NP-SBJ}(q_{\text{NML}}, q_V) \xrightarrow{\text{q}_{\text{NP-SBJ}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))
\]
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

\[
\begin{align*}
NML(q_Y, q_P) & \xrightarrow{q_{NML}} NP(q_P, q_Y) \\
NP(q_S) & \xrightarrow{q_{NP}} NP(q_S) \\
PP(q_f, q_{NP}) & \xrightarrow{q_{PP}} PP(q_f, q_{NP})
\end{align*}
\]

\[
\begin{align*}
NP-SBJ(q_{NML}, q_V) & \xrightarrow{q_{NP-SBJ}} NP-SBJ(q_{NML}, NP(q_V))
\end{align*}
\]
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

\[
\begin{align*}
\text{NML}(q_Y, q_P) & \xrightarrow{q_{\text{NML}}} \text{NP}(q_P, q_Y) \\
\text{NP}(q_S) & \xrightarrow{q_{\text{NP}}} \text{NP}(q_S) \\
\text{PP}(q_f, q_{\text{NP}}) & \xrightarrow{q_{\text{PP}}} \text{PP}(q_f, q_{\text{NP}}) \\
\text{NP-SBJ}(q_{\text{NML}}, q_V) & \xrightarrow{q_{\text{NP-SBJ}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))
\end{align*}
\]
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

\[
\text{NML}(q_Y, q_P) \xrightarrow{q_{\text{NML}}} \text{NP}(q_P, q_Y)
\]
\[
\text{NP}(q_S) \xrightarrow{q_{\text{NP}}} \text{NP}(q_S)
\]
\[
\text{PP}(q_f, q_{\text{NP}}) \xrightarrow{q_{\text{PP}}} \text{PP}(q_f, q_{\text{NP}})
\]

\[
\text{NP-SBJ}(q_{\text{NML}}, q_V) \xrightarrow{q_{\text{NP-SBJ}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))
\]
Syntax-based Machine Translation

- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

\[
\text{NML}(q_Y, q_P) \xrightarrow{q_{\text{NML}}} \text{NP}(q_P, q_Y) \\
\text{NP}(q_S) \xrightarrow{q_{\text{NP}}} \text{NP}(q_S) \\
\text{PP}(q_f, q_{\text{NP}}) \xrightarrow{q_{\text{PP}}} \text{PP}(q_f, q_{\text{NP}})
\]

\[
\text{NP-SBJ}(q_{\text{NML}}, q_V) \xrightarrow{q_{\text{NP-SBJ}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))
\]

Rule Extraction for MT
# Syntax-based Machine Translation

## Rules

<table>
<thead>
<tr>
<th>Yugoslav $q_Y$ AlywgwslAf</th>
<th>President $q_P$ AlywslAf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voislav $q_V$ fwyslAf</td>
<td>President for $q_f$ En</td>
</tr>
<tr>
<td>Serbia $q_S$ SrbyA</td>
<td></td>
</tr>
</tbody>
</table>

- NML($q_Y$, $q_P$) $\xrightarrow{q_{NML}}$ NP($q_P$, $q_Y$)
- PP($q_f$, $q_{NP}$) $\xrightarrow{q_{PP}}$ PP($q_f$, $q_{NP}$)
- NP-SBJ($q_{NML}$, $q_V$) $\xrightarrow{q_{NP-SBJ}}$ NP-SBJ($q_{NML}$, NP($q_V$))
Syntax-based Machine Translation

<table>
<thead>
<tr>
<th>Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yugoslav</strong> $^{q_Y}$ AlywgwsIAfy</td>
</tr>
<tr>
<td><strong>Voislav</strong> $^{q_V}$ fwysIAf</td>
</tr>
<tr>
<td><strong>Serbia</strong> $^{q_S}$ SrbyA</td>
</tr>
</tbody>
</table>

**NML**($q_Y$, $q_P$) $^{q_{NML}}$ **NP**($q_P$, $q_Y$)

**PP**($q_f$, $q_{NP}$) $^{q_{PP}}$ **PP**($q_f$, $q_{NP}$)

**NP-SBJ**($q_{NML}$, $q_V$) $^{q_{NP-SBJ}}$ **NP-SBJ**($q_{NML}$, **NP**($q_V$))

→ Rules of an Extended Top-down Tree Transducer
Extended Top-down Tree Transducer

**Advantages**

- simple and natural model
- easy to train (from linguistic resources) [GRAEHL et al., 2008]
- symmetric
Extended Top-down Tree Transducer

Advantages

- simple and natural model
- easy to train (from linguistic resources) [GRAEHL et al., 2008]
- symmetric

Disadvantages

- weights set in the same manner
- does not capture utility, EM available
  → EM not used in practice
corresponds to two productions

\[
q_{\text{NP-SBJ}} \rightarrow \text{NP-SBJ}(q_{\text{NML}}, q_{\text{V}})
\]

\[
q_{\text{NP-SBJ}} \rightarrow \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_{\text{V}}))
\]
General idea

Synchronous grammars are essentially two grammars over the same nonterminals whose productions are paired.

Convention

same nonterminals are synchronized (or linked) and develop at the same time.
Approach

- join two productions $q_1 \rightarrow r_1$ and $q_2 \rightarrow r_2$ to 
(\(q_1, q_2\) \(\rightarrow\) \(r_1, r_2\))
join two productions $q_1 \rightarrow r_1$ and $q_2 \rightarrow r_2$ to
$(q_1, q_2) \rightarrow (r_1, r_2)$

demand $q_1 = q = q_2$ for simplicity and write $r_1 \xrightarrow{q} r_2$
From Automata to Transducers

Approach

- join two productions $q_1 \rightarrow r_1$ and $q_2 \rightarrow r_2$ to $(q_1, q_2) \rightarrow (r_1, r_2)$
- demand $q_1 = q = q_2$ for simplicity and write $r_1 \xrightarrow{q} r_2$
- paired productions develop input and output tree at the same time
From Automata to Transducers

Used rule:  

Next rule:  

\[ q \rightarrow S \text{CONJ} wa q \]
From Automata to Transducers

Used rule:

```
  q
/\  
S CONJ q
```

```
q
S
CONJ
wa
```

Next rule:

```
  q
/\  
S CONJ q
```

```
  q
S
CONJ
wa
```

```
q
S
VP
q1
p
q2
```

```
q
S
p
q1
q2
```
From Automata to Transducers

Used rule:

Next rule:

Rule Extraction for MT
From Automata to Transducers

Used rule:

Next rule:

Rule Extraction for MT
From Automata to Transducers

Used rule:

Next rule:
From Automata to Transducers

Used rule:

```
N  r  N
boy atefl
```

Next rule:

```
NP  r  q2  NP
```

Rule Extraction for MT
From Automata to Transducers

Used rule:

Next rule:

Rule Extraction for MT
From Automata to Transducers

Used rule:  

\[
\text{door} \rightarrow r \quad \text{albab}
\]

Next rule:

\[
\text{wa} \quad \text{ra’aa} \quad \text{atefl} \quad \text{albab}
\]
Remarks

- synchronization breaks almost all existing constructions (e.g., the normalization construction)
- the basic grammar model very important
### Other Syntax-based Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SCFG</strong> = synchronous context-free grammar</td>
<td>LTG-LTG (synchronous local tree grammar)</td>
</tr>
<tr>
<td>[Chiang, 2007]</td>
<td>⊆ ln-TOP</td>
</tr>
<tr>
<td><strong>STSG</strong> = synchronous tree substitution grammar</td>
<td>TSG-TSG (special extended top-down tree transducer)</td>
</tr>
<tr>
<td>[Eisner, 2003]</td>
<td>⊆ ln-XTOP</td>
</tr>
<tr>
<td><strong>STAG</strong> = synchronous tree adjunction grammar</td>
<td>TAG-TAG (special extended top-down tree transducer)</td>
</tr>
<tr>
<td>[Shieber, Schabes, 1990]</td>
<td></td>
</tr>
<tr>
<td><strong>SCFTG</strong> = synchronous context-free tree grammar</td>
<td>CFTG-CFTG</td>
</tr>
<tr>
<td>[Nederhof, Vogler, 2012]</td>
<td></td>
</tr>
</tbody>
</table>
Other Syntax-based Models

Towards asymmetric relations

- **STSSG** = synch. tree-sequence substitution grammar
  \[ \text{ZHANG et al., 2008} \]
  \[ \text{TSSG-TSSG} \]

- **$\ell$MBOT** = local shallow multi bottom-up tree transducer
  \[ \text{BRAUNE et al., 2013} \]
  \[ \text{LTG-TSSG} \]

In-$X$MBOT corresponds roughly to \text{RTG-TSSG}
Where is Machine Learning?

<table>
<thead>
<tr>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>- unsupervised learning of translation models for STSG</td>
</tr>
<tr>
<td>[BLUNSON et al., 2008]</td>
</tr>
<tr>
<td>- uses GIBBS sampling, hierarchical DIRICHLET process, ...</td>
</tr>
</tbody>
</table>
Where is Machine Learning?

Examples
- unsupervised learning of translation models for STSG [Blunsom et al., 2008]
- uses Gibbs sampling, hierarchical Dirichlet process, …

But …
- can only be used on small data
- does not scale well
→ currently not used in practice
Quo vadis?
Selected references

**ARNOLD, DAUCHET**: *Morphismes et Bimorphismes d’Arbres*  

**BRAUNE, SEEMANN, QUERNHEIM, ~**: *Shallow local multi bottom-up tree transducers in statistical machine translation.*  
Proc. ACL, 2013

**GALLEY, HOPKINS, KNIGHT, MARCU**  
*What’s in a Translation Rule?*  
Proc. NAACL, 2004

**KOEHN**: *Statistical Machine Translation*  
Cambridge University Press, 2010

**~, GRAEHL, HOPKINS, KNIGHT**: *The Power of Extended Top-down Tree Transducers.*  

**OCH, NEY**: *The Alignment Template Approach to Statistical Machine Translation.*  
Comput. Linguist. 30, 2004