Expressive Power of Syntax-based Machine Translation Formalisms

Andreas Maletti

Institute for Natural Language Processing
Universität Stuttgart

Stuttgart — December 8, 2015
1. The room it is not narrowly was a simple, bathtub was also attached.
2. Wi-fi, TV and I was available.
3. Church looked When morning awake open the curtain.
4. When looking at often, wives, went out and is invited to try to go […].
5. But was a little cold, morning walks was good.
The room it is not narrowly was a simple, bathtub was also attached.

Wi-fi, TV and I was available.

Church looked When morning awake open the curtain.

When looking at often, wives, went out and is invited to try to go [...].

But was a little cold, morning walks was good.
Okay, what is your name?
Abdul.
And your last name?
Al Farran.
Okay, what is your name?
Abdul.
And your last name?
Al Farran.

Okay, what’s your name?
milk a mechanic and I am here I mean yes
### Speech transcript

Okay, what is your name?

Abdul.

And your last name?

Al Farran.

### Speech translation [undisclosed system; most likely phrase-based]

Okay, what’s your name?

milk a mechanic and I am here I mean yes

What is your last name?

every two weeks my son’s name is ismail
Short History

Timeline

1960 - Dark age
- rule-based systems (e.g., SYSTRAN)
- Chomskyan approach (perfect translation, poor coverage)

1991 - Reformation
- phrase-based and syntax-based systems
- statistical approach (cheap, automatically trained)

2015 - Potential future
- semantics-based systems (e.g., FrameNet-based)
- semi-supervised, statistical approach
- basic understanding of (translated) text
Vauquois triangle:

Translation model:
Vauquois triangle:

Translation model: *string-to-tree*
Machine Translation

Vauquois triangle:

Translation model: *tree-to-tree*
Machine Translation

Training data

- parallel corpus
- word alignments
- parse trees
Training data
- parallel corpus
- word alignments
- parse trees

Parallel Corpus
linguistic resource containing *(sentence-by-sentence)* example translations
I would like your advice about Rule 143 concerning inadmissibility.

Könnten Sie mir eine Auskunft zu Artikel 143 im Zusammenhang mit der Unzulässigkeit geben.
parallel corpus, *word alignments*, parse tree

I would like your advice about Rule 143 concerning inadmissibility.

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

Synchronous Grammars

Synchronous tree substitution grammar: productions \( N \rightarrow (r, r_1) \)

- nonterminal \( N \)
- right-hand side \( r \) of context-free grammar production
- right-hand side \( r_1 \) of tree substitution grammar production

Synchronous Grammars

Synchronous tree substitution grammar: productions $N \rightarrow (r, r_1)$

- nonterminal $N$
- right-hand side $r$ of context-free grammar production
- right-hand side $r_1$ of tree substitution grammar production
- (bijective) synchronization of nonterminals

Synchronous Grammars

Production application

1 Selection of synchronous nonterminals
Synchronous Grammars

Production application

1. Selection of synchronous nonterminals
Production application

1. Selection of synchronous nonterminals
2. Selection of suitable production

Synchronous Grammars

S → PPER KOUS PPER would like KOUS PPER Radvice PP

Könnten
eine Auskunft geben

S → KOUS PPER PPER ART NN PP VV

Expressive Power of Syntax-based Machine Translation Formalisms

Andreas Maletti
Synchronous Grammars

Production application

1. Selection of synchronous nonterminals
2. Selection of suitable production
3. Replacement on both sides
**Synchronous Grammars**

Production application

1. synchronous nonterminals
Synchronous Grammars

Production application

**synchronous nonterminals**

Könnten eine Auskunft geben

Synchronous nonterminals

Expressive Power of Syntax-based Machine Translation Formalisms
Synchronous Grammars

Production application

1. synchronous nonterminals
2. suitable production
Production application

1. synchronous nonterminals
2. suitable production
3. replacement
I would like your advice about Rule 143 concerning inadmissibility.

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

would like your advice about Rule 143 concerning inadmissibility

I would like your advice about Rule 143 concerning inadmissibility.


Expressive Power of Syntax-based Machine Translation Formalisms

Andreas Maletti
I would like your advice about Rule 143 concerning inadmissibility following [Galley, Hopkins, Knight, Marcu: What’s in a translation rule? Proc. NAACL, 2004]
I would like your advice about Rule 143 concerning inadmissibility following [Galley, Hopkins, Knight, Marcu: What’s in a translation rule? Proc. NAACL, 2004]
Removal of extractable production:

I would like your advice about Rule 143 concerning inadmissibility.
Production Extraction

Removal of extractable production:

Könnten Sie eine Auskunft zu Artikel 143 geben?
Repeate production extraction:

Könnten Sie eine Auskunft zu Artikel 143 geben.

Andreas Maletti
PPER would like your advice about Rule 143.
Könnten Sie eine Auskunft zu Artikel 143 geben?
Repeat production extraction:

KÖNNTEN SIE EINE AUSKUNFT ZU ARTIKEL 143 GEBEN?

Would you like your advice about Rule 143?
Repetitive production extraction:

(Extractable productions marked in red)
PPER would like your advice about Rule 143.

Könnten Sie eine Auskunft zu Artikel 143 geben?
Synchronous Tree Substitution Grammars

Advantages

- very simple
- implemented in framework ‘Moses’
- “context-free”
Synchronous Tree Substitution Grammars

Advantages

- very simple
- implemented in framework ‘Moses’
- “context-free”

Disadvantages

- problems with discontinuities
- composition and binarization not possible
- “context-free”
### Evaluation

**English → German translation task:** (higher BLEU is better)

<table>
<thead>
<tr>
<th>Type</th>
<th>System</th>
<th>BLEU [vanilla]</th>
<th>BLEU [competition]</th>
</tr>
</thead>
<tbody>
<tr>
<td>string-to-string</td>
<td>phrase-based</td>
<td>16.7</td>
<td>20.3</td>
</tr>
<tr>
<td></td>
<td>hierarchical</td>
<td>17.0</td>
<td>—</td>
</tr>
<tr>
<td>string-to-tree</td>
<td>STSG</td>
<td>15.2</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>STSG [GHKM]</td>
<td>17.1</td>
<td>19.4</td>
</tr>
<tr>
<td>tree-to-tree</td>
<td>STSG</td>
<td>14.5</td>
<td>—</td>
</tr>
</tbody>
</table>

Overview

1. Background

2. Extending the Expressive Power

3. Investigating their Expressive Power
very specific production

every production for ‘advice’ contains sentence structure

(syntax “in the way”)
Synchronous Grammars

Synchronous multi tree substitution grammar: $N \rightarrow (r, \langle r_1, \ldots, r_n \rangle)$


- nonterminal $N$
- right-hand side $r$ of context-free grammar production
- right-hand sides $r_1, \ldots, r_n$ of regular tree grammar production
Synchronous Grammars

Synchronous multi tree substitution grammar: $N \rightarrow (r, \langle r_1, \ldots, r_n \rangle)$


- nonterminal $N$
- right-hand side $r$ of context-free grammar production
- right-hand sides $r_1, \ldots, r_n$ of regular tree grammar production
Synchronous multi tree substitution grammar: $N \rightarrow (r, \langle r_1, \ldots, r_n \rangle)$

- nonterminal $N$
- right-hand side $r$ of context-free grammar production
- right-hand sides $r_1, \ldots, r_n$ of regular tree grammar production

Synchronization via map $NT$ $\rightarrow$ $r_1, \ldots, r_n$ to $NT$

ART-NN-VV $\rightarrow$

<table>
<thead>
<tr>
<th>( \text{ART} )</th>
<th>( \text{NN} )</th>
<th>( \text{VV} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{eine} )</td>
<td>( \text{Auskunft} )</td>
<td>( \text{geben} )</td>
</tr>
</tbody>
</table>
Synchronous Grammars

Synchronous multi tree substitution grammar: \( N \rightarrow (r, \langle r_1, \ldots, r_n \rangle) \)

- nonterminal \( N \)
- right-hand side \( r \) of context-free grammar production
- right-hand sides \( r_1, \ldots, r_n \) of regular tree grammar production

\[ \text{ART-NN-VV} \rightarrow \text{NP-VV} \]

ART NN APPR NN CD PP VV

NP-VV \( \rightarrow \)

PP

zu Artikel 143
Synchronous Grammars

Synchronous multi tree substitution grammar: $N \rightarrow (r, \langle r_1, \ldots, r_n \rangle)$

- nonterminal $N$
- right-hand side $r$ of context-free grammar production
- right-hand sides $r_1, \ldots, r_n$ of regular tree grammar production
- synchronization via map NT $r_1, \ldots, r_n$ to NT $r$

NP -VV → ART-NN-VV about Rule 143 PP

zu Artikel 143 CD PP VV

NP PP

ART NN APPR NN CD PP VV

Expressive Power of Syntax-based Machine Translation Formalisms
Production application

synchronous nonterminals
Synchronous Grammars

Production application

1 synchronous nonterminals
Synchronous Grammars

Production application

1. synchronous nonterminals
2. suitable production

ART-NN-VV →

NP-VV →

ART NN APPR NN CD PP VV

NP Artikel 143 PP

zu Artikel 143

ART-NN-VV about Rule 143 PP

ADVICE...
Synchronous Grammars

Production application

1. synchronous nonterminals
2. suitable production
3. replacement
PPER would like your advice about Rule 143.

Könnten Sie eine Auskunft zu Artikel 143 geben?

variant of [M.: How to train your multi bottom-up tree transducer. Proc. ACL, 2011]
PPER would like your advice about Rule 143.

variant of [M.: How to train your multi bottom-up tree transducer. Proc. ACL, 2011]
PPER would like your advice about Rule 143

variant of [M.: How to train your multi bottom-up tree transducer. Proc. ACL, 2011]
PPER would like your advice about Rule 143.

Könnten Sie eine Auskunft zu Artikel 143 geben?

variant of [M.: How to train your multi bottom-up tree transducer. Proc. ACL, 2011]
Advantages

- complicated discontinuities
- implemented in framework ‘Moses’
  
- binarizable, composable
Synchronous Multi Tree Substitution Grammars

Advantages

- complicated discontinuities
- implemented in framework ‘Moses’  
- binarizable, composable

Disadvantages

- output non-regular (trees) or non-context-free (strings)
- not symmetric (input context-free; output not)
He bought a new and fuel-efficient car.
<table>
<thead>
<tr>
<th>Task</th>
<th>System</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>English → German</td>
<td>STSG</td>
<td>15.2</td>
</tr>
<tr>
<td></td>
<td>SMTSG</td>
<td>*15.9</td>
</tr>
<tr>
<td>English → Arabic</td>
<td>STSG</td>
<td>48.3</td>
</tr>
<tr>
<td></td>
<td>SMTSG</td>
<td>*49.1</td>
</tr>
<tr>
<td>English → Chinese</td>
<td>STSG</td>
<td>17.7</td>
</tr>
<tr>
<td></td>
<td>SMTSG</td>
<td>*18.4</td>
</tr>
<tr>
<td>English → Polish</td>
<td>STSG</td>
<td>21.3</td>
</tr>
<tr>
<td></td>
<td>SMTSG</td>
<td>*23.4</td>
</tr>
<tr>
<td>English → Russian</td>
<td>STSG</td>
<td>24.7</td>
</tr>
<tr>
<td></td>
<td>SMTSG</td>
<td>*26.1</td>
</tr>
</tbody>
</table>

Overview

1. Background

2. Extending the Expressive Power

3. Investigating their Expressive Power
Synchronous Grammars

Notes
- tree-to-tree models easier for theoretical investigation
- strongly related to tree transducers
- we disallow trivial input sides of just a nonterminal (ε-free)

Synchronous grammar:

Tree transducer:
Synchronous Grammars

Major linear tree transducers:

<table>
<thead>
<tr>
<th>synchronization</th>
<th>bijective</th>
<th>injective</th>
</tr>
</thead>
<tbody>
<tr>
<td>input sides</td>
<td>(output → input)</td>
<td></td>
</tr>
<tr>
<td>shallow</td>
<td>nondeleting top-down …</td>
<td>top-down …</td>
</tr>
<tr>
<td>general</td>
<td>nondeleting extended …</td>
<td>extended …</td>
</tr>
</tbody>
</table>

Further distinction

- allow productions on disconnected input nonterminals
  → regular look-ahead
- allow arbitrary trees for disconnected input nonterminals
  → no look-ahead
Synchronous Grammars

Illustration

- **no look-ahead**: can plug any (terminal) tree for $N_{MD}$
- [e.g., $NP(DT(\text{the}), NN(\text{tower}))$]

Expressive Power of Syntax-based Machine Translation Formalisms

Andreas Maletti
Synchronous Grammars

Illustration

- **no look-ahead**: can plug any (terminal) tree for $N_{MD}$
  
  [e.g., $NP(DT(\text{the}), NN(\text{tower}))$]

- **regular look-ahead**: use special “no-output”-productions $N \rightarrow (r)$
  
  [e.g., $N_{MD} \rightarrow (MD(\text{should}))$]
Synchronous Grammars

Illustration

- **no look-ahead**: can plug any (terminal) tree for $N_{MD}$
  
  \[e.g., NP(\text{DT}(\text{the}), \text{NN}(\text{tower}))\]

- **regular look-ahead**: use special “no-output”-productions $N \rightarrow (r)$
  
  \[e.g., N_{MD} \rightarrow (\text{MD}(\text{should}))\]

- SMTSG always have regular look-ahead
  (any number of components includes 0)
Synchronous Grammars

Evaluation Criteria

rotations implementable? (for arbitrary $t_1, t_2, t_3$)

symmetric?
domain regular?
range regular?
closed under composition?


Icons by interactivemania (http://www.interactivemania.com/) and UN Office for the Coordination of Humanitarian Affairs
Synchronous Grammars

Illustration of rotations

- The diagram shows two syntactic structures, illustrating how the grammars can be rotated.
- The left side of the diagram has a structure with 'Alice' and 'carries' as a NP and VBD, respectively, under a VP node.
- The right side of the diagram has a structure with 'Bob' and 'carried' as a NP and VBN, respectively, under a VP node.

Expressive Power of Syntax-based Machine Translation Formalisms
Andreas Maletti · 29
Hasse diagram with composition closure indicated in subscript:
## Top-down Tree Transducer

| Model \ Criterion | \(\xmark\) | \(\checkmark\) | \(\xmark\) | \(\checkmark\) |
|-------------------|-------------|----------------|-------------|
| ns-TOP            | \(\xmark\) | \(\checkmark\) | \(\xmark\) | \(\checkmark\) |
| n-TOP             | \(\xmark\) | \(\checkmark\) | \(\xmark\) | \(\checkmark\) |
| s-TOP             | \(\xmark\) | \(\checkmark\) | \(\checkmark\) | \(\checkmark\) |
| s-TOP\(^R\)       | \(\xmark\) | \(\checkmark\) | \(\checkmark\) | \(\checkmark\) |
| TOP               | \(\xmark\) | \(\checkmark\) | \(\checkmark\) | \(\xmark\) |
| TOP\(^R\)         | \(\xmark\) | \(\checkmark\) | \(\checkmark\) | \(\checkmark\) |
Hasse diagram with the composition closure indicated in subscript:

composition closures by
### Synchronous Tree Substitution Grammars

<table>
<thead>
<tr>
<th>Model \ Criterion</th>
<th>$\text{TOP}$</th>
<th>$\text{TOP}^R$</th>
<th>$\text{ns-STSG}$</th>
<th>$\text{n-STSG}$</th>
<th>$\text{s-STSG}^{(R)}$</th>
<th>$\text{STSG}$</th>
<th>$\text{STSG}^R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-TOP</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>TOP</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X₂</td>
</tr>
<tr>
<td>TOP$^R$</td>
<td>X</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ns-STSG</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X₂</td>
<td>✓</td>
</tr>
<tr>
<td>n-STSG</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X∞</td>
<td>X₂</td>
</tr>
<tr>
<td>s-STSG$^{(R)}$</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X₂</td>
<td>✓</td>
</tr>
<tr>
<td>STSG</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X₄</td>
<td>X₃</td>
</tr>
<tr>
<td>STSG$^R$</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X₃</td>
<td></td>
</tr>
</tbody>
</table>
Synchronous Multi Tree Substitution Grammars

Advantages of SMTSG

- always have regular look-ahead
- can always be made nondeleting & shallow
- closed under composition

Synchronous Multi Tree Substitution Grammars

Advantages of SMTSG

- always have regular look-ahead
- can always be made nondeleting & shallow
- closed under composition

Disadvantages of SMTSG

- non-regular range

Synchronous Multi Tree Substitution Grammars

Hasse diagram with the composition closure indicated in subscript:
### Synchronous Multi Tree Substitution Grammars

<table>
<thead>
<tr>
<th>Model \ Criterion</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-TOP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>TOP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X₂</td>
</tr>
<tr>
<td>TOP&lt;sup&gt;R&lt;/sup&gt;</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ns-STSG</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>X₂</td>
</tr>
<tr>
<td>n-STSG</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>X&lt;sub&gt;∞&lt;/sub&gt;</td>
</tr>
<tr>
<td>s-STSG&lt;sup&gt;(R)&lt;/sup&gt;</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>X₂</td>
</tr>
<tr>
<td>STSG</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>X₄</td>
</tr>
<tr>
<td>STSG&lt;sup&gt;R&lt;/sup&gt;</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>X₃</td>
</tr>
<tr>
<td>(n)s-SMTSG&lt;sup&gt;(R)&lt;/sup&gt;</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>(n)-SMTSG&lt;sup&gt;(R)&lt;/sup&gt;</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>reg.-range SMTSG</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>symmetric SMTSG</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>


---

**Expressive Power of Syntax-based Machine Translation Formalisms**

Andreas Maletti
Synchronous Multi Tree Substitution Grammars

Theorem

$$\text{SMTSG}^3 \preceq \text{reg.-range SMTSG}$$

Synchronous Multi Tree Substitution Grammars

Counterexample relation

- abstracts a well-known linguistic transformation called **topicalization**
- implementable by SMTSG, but not by any composition of STSG
Illustration of topicalization

- It rained **yesterday night**.

  **Topicalized:** **Yesterday night**, it rained.
Illustration of topicalization

- It rained yesterday night.
  Topicalized: Yesterday night, it rained.

- We toiled all day yesterday at the restaurant that charges extra for clean plates.
  Topicalized: At the restaurant that charges extra for clean plates, we toiled all day yesterday.
On the tree level

- **S**
  - **NP**
    - **PRP** we
    - **VBD** toiled
  - **NP**
    - **DT** all
    - **NN** day
    - **NN** yesterday
  - **PP**
    - **IN** at
    - **NP**
      - **NN** the
      - **NN** restaurant
    - **SBAR**
      - **WHNP** that
      - **S**
        - **VP**
          - **VBZ** charges
          - **NP**
            - **JJ** extra
            - **PP**
              - **IN** for
              - **JJ** NNS
                - **clean**
                - **NNS** plates

- **S**
  - **PP**
    - **IN** at
    - **NP**
      - **DT** the
      - **NN** restaurant
    - **SBAR**
      - **WHNP** that
      - **S**
        - **VP**
          - **VBZ** charges
          - **NP**
            - **JJ** extra
            - **PP**
              - **IN** for
              - **JJ** NNS
                - **clean**
                - **NNS** plates
Summary

Contributions

- **SMTSG implementation and evaluation**
  - [Braune, Seemann, Quernheim, M.: Shallow local multi bottom-up tree transducers in SMT. *Proc. ACL*, 2013]

- **characterization of expressive power of STSG and SMTSG**

- **new proof technique (based on synchronization links)**
  - [Filiot, Maneth, Reynier, Talbot: Decision problems of tree transducers with origin. *Proc. ICALP*, 2015]
Contributions

- **SMTSG implementation and evaluation**
  
  [Braune, Seemann, Quernheim, M.: Shallow local multi bottom-up tree transducers in SMT. *Proc. ACL*, 2013]

- **characterization of expressive power of STSG and SMTSG**


- **new proof technique** (based on synchronization links)

  similar ideas used in
  [Filiot, Maneth, Reynier, Talbot: Decision problems of tree transducers with origin. *Proc. ICALP*, 2015]

Thank you for the attention.