Using Linguistic Syntax in Statistical Machine Translation

Greg Hanneman

11-731: Machine Translation
March 22, 2016
Syntax in Machine Translation

What does that even mean?
Syntax in Machine Translation

What does that even mean?

les voitures bleues

I like books that children read.

J’aime que les fleurs poussent dans les champs.

The final vote will take place tomorrow at noon.

I like books that children read.

Le vote final aura lieu.
Syntax in Machine Translation

Why would we want that?
Syntax in Machine Translation

Why would we want that?

• Language is hierarchical, not finite-state!
• Generalize over different/unseen strings with same structure
• Languages have structural divergences in addition to lexical ones
• More accurate reordering at arbitrary distances
Syntax in Machine Translation

How would we make it happen?
Syntax in Machine Translation

How would we make it happen?

- What notion of syntax?
- Syntax on source, target, or both?
- What do the rules look like?
- What do the labels look like?
- Input is string, tree, or forest?
- Decode bottom-up or top-down?
# Syntax in Machine Translation

<table>
<thead>
<tr>
<th>Side(s)</th>
<th>Rules</th>
<th>Labels</th>
<th>Input</th>
<th>Technique(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>SCFG</td>
<td>Treebank</td>
<td>Tree</td>
<td>Yamada and Knight [2001]</td>
</tr>
<tr>
<td>Source</td>
<td>STSG</td>
<td>Treebank</td>
<td>Tree</td>
<td>Huang et al. [2006], Liu et al. [2006]</td>
</tr>
<tr>
<td>Source</td>
<td>STSG</td>
<td>Treebank</td>
<td>Forest</td>
<td>Mi et al. [2008]</td>
</tr>
<tr>
<td>Target</td>
<td>SCFG</td>
<td>Treebank + combinations</td>
<td>String</td>
<td>Zollmann &amp; Venugopal [2006]</td>
</tr>
<tr>
<td>Target</td>
<td>SCFG</td>
<td>“X” with soft preferences</td>
<td>String</td>
<td>Venugopal et al. [2009]</td>
</tr>
<tr>
<td>Target</td>
<td>STSG</td>
<td>Treebank</td>
<td>String</td>
<td>Galley et al. [2004]</td>
</tr>
<tr>
<td>Target</td>
<td>SCFG</td>
<td>Syntactic “X” with soft prefs</td>
<td>String</td>
<td>Huck et al. [2014]</td>
</tr>
<tr>
<td>Both</td>
<td>STSG</td>
<td>Bi-treebank</td>
<td>Tree</td>
<td>Xia and McCord [2004]</td>
</tr>
<tr>
<td>Both</td>
<td>SCFG</td>
<td>Bi-treebank</td>
<td>String</td>
<td>Lavie et al. [2008]</td>
</tr>
<tr>
<td>Both</td>
<td>STSG</td>
<td>Bi-treebank</td>
<td>Forest</td>
<td>Liu et al. [2009]</td>
</tr>
<tr>
<td>Both</td>
<td>SCFG</td>
<td>Bi-treebank + virtual nodes</td>
<td>String</td>
<td>Hanneman et al. [2011]</td>
</tr>
<tr>
<td>Both</td>
<td>SCFG</td>
<td>Bi-treebank + combinations, with soft matches</td>
<td>String</td>
<td>Chiang [2010]</td>
</tr>
</tbody>
</table>
The final vote will take place tomorrow at noon.
The final vote will take place tomorrow at noon.

X → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]
Le vote final aura lieu demain à 12 heures.

X → [aura lieu demain à 12 heures]: [will take place tomorrow at noon]
X → [demain à 12 heures]: [tomorrow at noon]
The final vote will take place tomorrow at noon.

X → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]
X → [demain à 12 heures]::[tomorrow at noon]
X → [aura lieu X₁]::[will take place X₁]
Hiero Pros and Cons

- Synchronous context-free grammar (SCFG)
- Easy to get given basic phrase extraction
- Sometimes works quite well

\[ X \rightarrow [\text{que } X_1 \text{ les } X_2]::[\text{that } X_2 X_1] \]

J’aime les livres que lisent les enfants.
I like books that children read.
Hierro Pros and Cons

- Synchronous context-free grammar (SCFG)
- Easy to get given basic phrase extraction
- But sometimes doesn’t...

\[ X \rightarrow [\text{que } X_1 \text{ les } X_2]::[\text{that } X_2 X_1] \]

J’ aime les livres que lisent les enfants.
I like books that children read.

✓

J’ aime que les fleurs poussent dans les champs.
I like that fields flowers grow in.

✗
The Trouble with “X”s

- Nonterminal boundaries can be anywhere.
- Can extract rule in one context and apply it in another – all “X”s look the same.
- “Formally but not linguistically syntactic”
Part 1

String-to-Tree Methods
“Syntax-Augmented” MT
[Zollmann and Venugopal, 2006; Zollmann, 2011]

• Let’s label these Hiero rules!
• Replace “X” with labels derived from a constituency parse of the target side
• Why target?
• What about phrases that aren’t constituents? Do we lose coverage?
Aside: Statistical Parsers

- Sentences
- Parse Trees
- New Sentence
- Statistical Parser
- New Parse Tree
The final vote will take place tomorrow at noon.
The final vote will take place tomorrow at noon.
Le vote final aura lieu demain à 12 heures.

VP → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]
ADVP+PP → [demain à 12 heures]::[tomorrow at noon]
The final vote will take place tomorrow at noon.

VP → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]
ADVP+PP → [demain à 12 heures]::[tomorrow at noon]
VP/PP → [aura lieu demain]::[take place tomorrow]
The final vote will take place tomorrow at noon.

VP → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]
ADVP+PP → [demain à 12 heures]::[tomorrow at noon]
VP → [aura lieu ADVP+PP₁]::[will take place ADVP+PP₁]
SAMT Labeling

• Try, in order:
  – A  exact constituent match
  – A+B adjacent constituents (not nec. siblings)
  – A/B partial A “missing a B to its right”
  – A\B partial A “missing a B to its left”
  – X if all else fails

• Slash labels are from combinatory categorial grammar

• About 4000 unique labels for English!
Translating with SAMT

NNS → [enfants]::[children]
NP → [J’]::[I]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]
S → [S₁ SBAR₂ .]::[S₁ SBAR₂ .]
S\NP → [aime les livres]::[like books]
SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]
SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]
VBP → [lisent]::[read]
Translating with SAMT

NNS → [enfants]::[children]  S\NP → [aime les livres]::[like books]
NP → [J’]::[I]  SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]  SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]
S → [S₁ SBAR₂.]::[S₁ SBAR₂.]  VBP → [lisent]::[read]

J’ aime les livres que lisent les enfants .  I
Translating with SAMT

NNS → [enfants]::[children]
NP → [J’]::[I]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]
S → [S₁ SBAR₂ .]::[S₁ SBAR₂ .]
S\NP → [aime les livres]::[like books]
SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]
SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]
VBP → [lisent]::[read]

J’ aime les livres que lisent les enfants .
I like books
Translating with SAMT

NNS → [enfants]::[children]
NP → [J’]::[I]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]
S → [S₁ SBAR₂ ]::[S₁ SBAR₂ ]
S\NP → [aime les livres]::[like books]
SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]
SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]
VBP → [lisent]::[read]
Translating with SAMT

NNS → [enfants]::[children]
NP → [J’]::[I]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]
S → [S₁ SBAR₂ .]::[S₁ SBAR₂ .]
S\NP → [aime les livres]::[like books]
SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]
SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]
VBP → [lisent]::[read]
Translating with SAMT

NNS $\rightarrow$ [enfants]::[children]
NP $\rightarrow$ [J’]::[I]
S $\rightarrow$ [NP$_1$ S\NP$_2$]::[NP$_1$ S\NP$_2$]
S $\rightarrow$ [S$_1$ SBAR$_2$ .]::[S$_1$ SBAR$_2$ .]
S\NP $\rightarrow$ [aime les livres]::[like books]
SBAR $\rightarrow$ [que VBP$_1$ les NNS$_2$]::[that NNS$_2$ VBP$_1$]
SBAR $\rightarrow$ [que S/NP$_1$ les NNS$_2$]::[that S/NP$_1$ the NNS$_2$]
VBP $\rightarrow$ [lisent]::[read]

J’ aime les livres que lisent les enfants .
I like books that children read
Translating with SAMT

NNS → [enfants]::[children]
NP → [J’]::[I]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]
S → [S₁ SBAR₂]::[S₁ SBAR₂]

S\NP → [aime les livres]::[like books]
SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]
SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]
VBP → [lisent]::[read]
Translating with SAMT

NNS → [enfants]::[children]
NP → [J’]::[I]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]
S → [S₁ SBAR₂ ]:[S₁ SBAR₂ ]

S\NP → [aime les livres]::[like books]
SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]
SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]
VBP → [lisent]::[read]

J’ aime les livres que lisent les enfants .
I like books that children read .
Translating with SAMT

NNS → [enfants]::[children]
NP → [J’]::[I]
S → [NP$_1$ S\NP$_2$]::[NP$_1$ S\NP$_2$]
S → [S$_1$ SBAR$_2$ .]::[S$_1$ SBAR$_2$ .]
S\NP → [aime les livres]::[like books]
SBAR → [que VBP$_1$ les NNS$_2$]::[that NNS$_2$ VBP$_1$]
SBAR → [que S/NP$_1$ les NNS$_2$]::[that S/NP$_1$ the NNS$_2$]
VBP → [lisent]::[read]

J’ aime les livres que lisent les enfants .
I like books that children read .
The Importance of Labels

NNS → [enfants]:[children]
NP → [J’]:[I]
S → [NP₁ S\NP₂]:[NP₁ S\NP₂]
S → [S₁ SBAR₂ .]:[S₁ SBAR₂ .]
S\NP → [aime les livres]:[like books]
SBAR → [que VBP₁ les NNS₂]:[that NNS₂ VBP₁]
SBAR → [que S/NP₁ les NNS₂]:[that S/NP₁ the NNS₂]
VBP → [lisent]:[read]
VB → [lisent]:[read]

J’aime les livres que lisent les enfants .
The Importance of Labels

NNS → [enfants]::[children]
NP → [J’]::[I]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]
S → [S₁ SBAR₂ .]::[S₁ SBAR₂ .]
S\NP → [aime les livres]::[like books]
SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]
SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]
VB → [lisent]::[read]
The Importance of Labels

NNS $\rightarrow$ [enfants]::[children]
NP $\rightarrow$ [J']::[I]
S $\rightarrow$ [NP$_1$ S\NP$_2$]::[NP$_1$ S\NP$_2$]
S $\rightarrow$ [S$_1$ SBAR$_2$.]::[S$_1$ SBAR$_2$.]
S\NP $\rightarrow$ [aime les livres]::[like books]
SBAR $\rightarrow$ [que VBP$_1$ les NNS$_2$]::[that NNS$_2$ VBP$_1$]
SBAR $\rightarrow$ [que S/NP$_1$ les NNS$_2$]::[that S/NP$_1$ the NNS$_2$]
VB $\rightarrow$ [lisent]::[read]

VB $\neq$ VBP
No match!
The Importance of Labels

NNS → [enfants]::[children]
NP → [J’]::[I]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]
S → [S₁ SBAR₂ .]::[S₁ SBAR₂ .]

S\NP → [aime les livres]::[like books]
SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]
SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]
VB → [lisent]::[read]

How do we finish parsing and translating?
The Importance of Labels

- Label of constituent being parsed must exactly match right-hand side of rule
- If no match in grammar, need a fallback: **glue rules**
- Finish translation without “real” syntax

\[
\begin{align*}
S & \rightarrow [X_1]::[X_1] \\
S & \rightarrow [S_1 X_2]::[S_1 X_2]
\end{align*}
\]
The Importance of Labels

- Label of constituent being parsed must exactly match right-hand side of rule
- If no match in grammar, need a fallback: glue rules
- Finish translation without “real” syntax

\[
S \rightarrow [X_1]::[X_1] \quad S \rightarrow [S_1 X_2]::[S_1 X_2] \\
S \rightarrow [NP_1]::[NP_1] \quad S \rightarrow [S_1 JJ+NN_2]::[S_1 JJ+NN_2] \\
S \rightarrow [S_1 NP_2]::[S_1 NP_2] \quad S \rightarrow [VP\VBZ_1]::[VP\VBZ_1] \\
S \rightarrow [S_1 VP\VBZ_2]::[S_1 VP\VBZ_2]
\]
The Importance of Labels

- Label of constituent being parsed must exactly match right-hand side of rule
- If no match in grammar, need a fallback: glue rules
- Finish translation without “real” syntax

One pair for each nonterminal in the grammar!
Translating with Glue Rules

NNS → [enfants]::[children]

NP → [J’]::[I]

S → [NP₁ S\NP₂]::[NP₁ S\NP₂]

S → [S₁ SBAR₂ .]::[S₁ SBAR₂ .]

S \NP → [aime les livres]::[like books]

SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]

SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]

VB → [lisent]::[read]

S → [VB₁]::[VB₁]
Translating with Glue Rules

NNS → [enfants]:[children]  S\NP → [aime les livres]:[like books]
NP → [J’]:[I]  S\NP → [aime les livres]:[like books]
S → [NP S\NP]:[NP S\NP]  SBAR → [que VBP les NNS]:[that NNS VBP]
S → [S_1 SBAR_2 ]:[S_1 SBAR_2 ]  SBAR → [que S/NP les NNS_2 ]:[that S/NP the NNS_2]
VB → [lisent]:[read]

X → [les]:[les]

J’ aime les livres que lisent les enfants .  I like books  read les children
Translating with Glue Rules

NNS $\rightarrow$ [enfants][children]
NP $\rightarrow$ [J’][I]
S $\rightarrow$ [NP$_1$ S\NP$_2$][NP$_1$ S\NP$_2$]
S $\rightarrow$ [S$_1$ SBAR$_2$.][S$_1$ SBAR$_2$.]
S\NP $\rightarrow$ [aime les livres][like books]
SBAR $\rightarrow$ [que VBP$_1$ les NNS$_2$][that NNS$_2$ VBP$_1$]
SBAR $\rightarrow$ [que S/NP$_1$ les NNS$_2$][that S/NP$_1$ the NNS$_2$]
VB $\rightarrow$ [lisent][read]
Translating with Glue Rules

NNS → [enfants]::[children]
NP → [J’]::[I]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]
S → [S₁ SBAR₂]::[S₁ SBAR₂]
S → [NP₁ S\NP₂]::[NP₁ S\NP₂]
S → [S₁ NNS₂]::[S₁ NNS₂]

S\NP → [aime les livres]::[like books]
SBAR → [que VBP₁ les NNS₂]::[that NNS₂ VBP₁]
SBAR → [que S/NP₁ les NNS₂]::[that S/NP₁ the NNS₂]
VB → [lisent]::[read]

J’aime les livres que lisent les enfants.
I like books read les children
SAMT Summary

• Label Hiero rules with nonterminals derived from parse trees

• Decisions made:
  – Syntax on: Target side
  – Rule formalism: SCFG
  – Input: String
  – Label space: Treebank + combinations

• Extensions: Venugopal et al. [2009], Chiang [2010], Zollmann [2011]
SAMT Pros and Cons
SAMT Pros and Cons

- No coverage loss from phrase-based MT
- Distinguishes different syntactic contexts

- Very large label set = sparse rules
- Still may need to restrict rule shape (Hiero)
- Produces weird tree structure
SAMT vs. “Native” Trees

I like books that children read.
SAMT vs. “Native” Trees

Stanford Parser

S
   / \
  NP   VP
     / \     
    PRP VBP NP SBAR
       /     \        
      I like NNS IN S

SBAR
  /   \
NP   VP
  /     
books that

SAMT decoding

S
   / \
  NP\NP VP
     /     
    I like books that NNS VBP

SBAR
  /   \
NP   VBP
     /     
children  read
• Let’s use larger fragments of tree structure!
• Keep internal nodes in the rule
• Require exact constituent match, but relax Hiero/SAMT rule format constraints
• Needs generalization of SCFG: tree-to-string transducer rules ("xRs") or synchronous tree-substitution grammar
Galley, Hopkins, Knight, Marcu

[Galley et al., 2004; Galley et al., 2006]

Extracted SAMT rule (SCFG)

$$\text{SBAR} \rightarrow [\text{que VBP}_1 \text{ les NNS}_2]::$$

$$[\text{that NNS}_2 \text{ VBP}_1]$$
Galley, Hopkins, Knight, Marcu
[Galley et al., 2004; Galley et al., 2006]

Extracted SAMT rule (SCFG)

\[ \text{SBAR} \rightarrow [\text{que VBP}_1 \text{ les NNS}_2]:: \]
\[ [\text{that NNS}_2 \text{ VBP}_1] \]

Extracted GHKM rule (xRs)

\[ \text{que VBP}_1 \text{ les NNS}_2 \rightarrow \text{SBAR} \]
\[ \begin{array}{c}
\text{IN} \\
\text{that} \\
\text{NP} \\
\text{NNS}_2 \\
\text{VBP}_1 \\
\text{VP} \\
\text{S} \\
\end{array} \]
Galley, Hopkins, Knight, Marcu

[Galley et al., 2004; Galley et al., 2006]

Extracted SAMT rule (SCFG)

\[
\text{SBAR} \rightarrow \text{[que VBP}_1 \text{ les NNS}_2]\text{:: [that NNS}_2 \text{ VBP}_1]
\]

Extracted GHKM rule (xRs)

\[
\text{que VBP}_1 \text{ les NNS}_2 \rightarrow \text{SBAR(IN(that) S(NP(NNS}_2 \text{ VP(VBP}_1)))}
\]
Two extraction methods:
- Minimal: break tree apart into smallest possible fragments
- Composed: combine minimal fragments into larger ones that might overlap

What about phrases that aren’t constituents? Do we lose coverage?
The final vote will take place tom. at noon.

lieu → NN(place)

NN → [lieu]::[place]
Le vote final aura lieu demain à 12 heures.

lieu → NN(place)
demain → RB(tomorrow)
Le vote final aura lieu demain à 12 heures.

lieu → NN(place)

12 heures → NN(noon)
demain → RB(tomorrow)
à → IN(at)
The final vote will take place tomorrow at noon.

lieu → NN(place)
demain → RB(tomorrow)
à → IN(at)

12 heures → NN(noon)
$\text{NN}_1 \rightarrow \text{NP}(\text{NN}_1)$ x2
$\text{RB}_1 \rightarrow \text{ADVP}(\text{RB}_1)$
The final vote will take place tom. at noon.
The final vote will take place tom. at noon.

aura NP₁ ADVP₂ PP₃ → VP(MD(will) VP(VB(take) NP₁ ADVP₂ PP₃))
GHKM Composed Rules

- When extracting minimal rules from a tree, plug two or more of them together.

12 heures $\rightarrow$ NN(noon)

+ \[ \text{NN}_1 \rightarrow \text{NP(NN}_1) \]

12 heures $\rightarrow$ NP(NN(noon))
When extracting minimal rules from a tree, plug two or more of them together:

12 heures → NN(noon)

NN₁ → NP(NN₁)

à → IN(at)

+ IN₁ NP₂ → PP(IN₁ NP₂)

à 12 heures → PP(IN(at) NP(NN(noon)))
Consider the Following

- Where do unaligned words go?
- What happens at the top of the tree for a very long/flat sentence?
- Is the decomposition of the tree unique?
Consider the Following

- Where do unaligned words go?
  - Minimal: Attach at highest-level rule by fiat
  - Composed: Try attaching to all rules

- What happens at the top of the tree for a very long/flat sentence?
  - You get a very long (and likely sparse) S rule

- Is the decomposition of the tree unique?
  - Minimal: Yes, given unaligned words fiat
  - Composed: Definitely not!
Translating with GHKM

aime → VBP(like)
enfants → NNS(children)
J’ → PRP(I)
lisent → VBP(read)
livres → NNS(books)
que → IN(that)
. → .(.)

les NNS₁ → NP(NNS₁)
IN₁ S₂ → SBAR(IN₁ S₂)
NP₁ VP₂ .₃ → S(NP₁ VP₂ .₃)
PRP₁ → NP(PRP₁)
VBP₁ → VP(VBP₁)
VP₁ NP₂ → S(NP₂ VP₁)
VBP₁ NP₂ SBAR₃ → VP(VBP₁ NP₂ SBAR₃)
Translating with GHKM

aime → VBP(like)
enfants → NNS(children)
J’ → PRP(I)
lisent → VBP(read)
livres → NNS(books)
que → IN(that)
. → .(.)
les NNS₁ → NP(NNS₁)
IN₁ S₂ → SBAR(IN₁ S₂)
NP₁ VP₂ .₃ → S(NP₁ VP₂ .₃)
PRP₁ → NP(PRP₁)
VBP₁ → VP(VBP₁)
VP₁ NP₂ → S(NP₂ VP₁)

J’aime les livres que lisent les enfants.

VP₁ NP₂ SBAR₃ →
VP(VBP₁ NP₂ SBAR₃)
J'aime les livres que lisent les enfants.
I like les livres que lisent les enfants.
Translating with GHKM

aime → VBP(like)

enfants → NNS(children)

J’ → PRP(I)

lisent → VBP(read)

livres → NNS(books)

que → IN(that)

. → .(.)

les NNS₁ → NP(NNS₁)

IN₁ S₂ → SBAR(IN₁ S₂)

NP₁ VP₂ .₃ → S(NP₁ VP₂ .₃)

PRP₁ → NP(PRP₁)

VBP₁ → VP(VBP₁)

VP₁ NP₂ → S(NP₂ VP₁)

VBP₁ NP₂ SBAR₃ →

VP(VBP₁ NP₂ SBAR₃)

PRP VBP NNS IN VBP NNS .

I like les books that read les children .
Translating with GHKM

aime → VBP(like)

enfants → NNS(children)

J’ → PRP(I)

lisent → VBP(read)

livres → NNS(books)

que → IN(that)

. → .(.)
Translating with GHKM

aime → VBP(like)

enfants → NNS(children)

J’ → PRP(I)

lisent → VBP(read)

livres → NNS(books)

que → IN(that)

. → .(.)

les NNS₁ → NP(NNS₁)

IN₁ S₂ → SBAR(IN₁ S₂)

NP₁ VP₂ .₃ → S(NP₁ VP₂ .₃)

PRP₁ → NP(PRP₁)

VBP₁ → VP(VBP₁)

VP₁ NP₂ → S(NP₂ VP₁)

VBP₁ NP₂ SBAR₃ →

VP(VBP₁ NP₂ SBAR₃)
Translating with GHKM

aime → VBP(like)
enfants → NNS(children)
J’ → PRP(I)
lisent → VBP(read)
livres → NNS(books)
que → IN(that)
. → .(.)
les NNS₁ → NP(NNS₁)
IN₁ S₂ → SBAR(IN₁ S₂)
NP₁ VP₂ .₃ → S(NP₁ VP₂ .₃)
PRP₁ → NP(PRP₁)
VBP₁ → VP(VBP₁)
VP₁ NP₂ → S(NP₂ VP₁)
VBP₁ NP₂ SBAR₃ → VP(VBP₁ NP₂ SBAR₃)
Translating with GHKM

aime → VBP(like)
enfants → NNS(children)
J’ → PRP(I)
lisent → VBP(read)
livres → NNS(books)
que → IN(that)
. → .(.)
les NNS₁ → NP(NNS₁)
IN₁ S₂ → SBAR(IN₁ S₂)
NP₁ VP₂ .₃ → S(NP₁ VP₂ .₃)
PRP₁ → NP(PRP₁)
VBP₁ → VP(VBP₁)
VP₁ NP₂ → S(NP₂ VP₁)
VBP₁ NP₂ SBAR₃ → VP(VBP₁ NP₂ SBAR₃)
Translating with GHKM

aime → VBP(like)
enfants → NNS(children)
J’ → PRP(I)
lisent → VBP(read)
livres → NNS(books)
que → IN(that)
. → (.)
les NNS₁ → NP(NNS₁)
IN₁ S₂ → SBAR(IN₁ S₂)
NP₁ VP₂ .₃ → S(NP₁ VP₂ .₃)
PRP₁ → NP(PRP₁)
VBP₁ → VP(VBP₁)
VP₁ NP₂ → S(NP₂ VP₁)
VBP₁ NP₂ SBAR₃ → VP(VBP₁ NP₂ SBAR₃)
Translating with GHKM

aime → VBP(like)
enfants → NNS(children)
J’ → PRP(I)
lisent → VBP(read)
livres → NNS(books)
que → IN(that)
. → .(.)
les NNS₁ → NP(NNS₁)
IN₁ S₂ → SBAR(IN₁ S₂)
NP₁ VP₂ ·₃ → S(NP₁ VP₂ ·₃)
PRP₁ → NP(PRP₁)
VBP₁ → VP(VBP₁)
VP₁ NP₂ → S(NP₂ VP₁)
VBP₁ NP₂ SBAR₃ → VP(VBP₁ NP₂ SBAR₃)
GHKM Summary

• Extract chunks of syntactic structure; recursively transform string into tree

• Decisions made:
  – Syntax on: Target side
  – Rule formalism: String-to-tree transducer
  – Input: String
  – Label space: Treebank

• Extensions and analysis: DeNeefe et al. [2007], Huck et al. [2014]
GHKM Pros and Cons
GHKM Pros and Cons

• More “native-looking” trees (LMs, etc.)
• No restrictions on rule shape or size
• Smaller label set than SAMT
• Can need big rules at top of tree
• No coverage for non-constituents
• Unary chains permit many label changes
Tree-to-Tree Methods
Parse Trees on Both Sides

• Why add the source-side tree?

• Why could that also be a bad idea?
Parse Trees on Both Sides

- Why add the source-side tree?
  - Further disambiguate syntactic context
  - Could parse input and transform it: don’t have to guess constituent boundaries

- Why could that also be a bad idea?
  - Could increase rule sparsity (e.g. labels)
  - Could hurt coverage if required to exactly match constituents on both sides
  - Only works for source and target languages that both have parsers
The final vote will take place.

N::NN → [vote]::[vote]
Tree-to-Tree Rule Extraction

N::NN → [vote]::[vote]
D::DT → [Le]::[The]
A::JJ → [final]::[final]
N::NN → [lieu]::[place]
PU::. → [.]::[.]
The final vote will take place.
Tree-to-Tree Methods

• Input is a string to jointly parse and translate (SCFG)  [Lavie et al., 2008]

• Input is a tree to translate via tree transduction (STSG)  [Xia and McCord, 2004]

• Input is a forest of packed variant trees to translate via transduction  [Liu et al., 2009]
Tree-to-Tree Relaxations

- Use SAMT labeling instead [Chiang, 2010]
- Insert “virtual” nodes into the tree to cover partial constituents [Ambati and Lavie, 2008; Hanneman et al., 2011]
- Allow rules to replace one subtree with multiple (STSSG) [Zheng et al., 2008]
- Or ignore the target-side tree entirely...
Part 3

Tree-to-String Methods
Tree-to-String Translation

- Extract chunks of syntactic structure; recursively transform tree into string
- Why model only the source tree?

- Do we need a new rule formalism?
Tree-to-String Translation

- Extract chunks of syntactic structure; recursively transform tree into string

- Why model only the source tree?
  - Take tree input at run time: don’t have to guess constituent bounds
  - Therefore very fast and no pruning!

- Do we need a new rule formalism?
  - Still STSG: just flip direction of GHKM-like rules
  - Could also use SCFG
Tree-to-String Methods

- Input is a tree to translate via transduction
  [Huang et al., 2006; Liu et al., 2006]

- Input is a tree where nodes can be inserted, permuted, or translated
  [Yamada and Knight, 2001]

- Input is a forest of packed variant trees to translate via transduction
  [Mi et al., 2008]

- Analysis: Neubig and Duh [2014]
Grand Summary
You Should Now Know...

- Why (linguistic) syntax in MT is desirable
- Some of the decision points involved
- Two ways of deciding: SAMT and GHKM
- How to extract/apply SCFG and STSG rules
- About trade-offs involving labels and tree constraints
- That if you want to know more...
There are Lots of Good Papers

- DeNeefe et al. (2007), “What Can Syntax-Based MT Learn from Phrase-Based MT?” Proceedings of EMNLP.
There are Lots of Good Papers

- Mi et al. (2008), “Forest-Based Translation.” Proceedings of ACL.
- Zhang et al. (2008), “A Tree Sequence Alignment-Based Tree-to-Tree Translation Model.” Proceedings of ACL.

Any questions?