A General-Purpose Rule Extractor for SCFG-Based Machine Translation

Greg Hanneman, Michelle Burroughs, and Alon Lavie

Language Technologies Institute
Carnegie Mellon University

Fifth Workshop on Syntax and Structure in Statistical Translation
June 23, 2011
SCFG Grammar Extraction

• Inputs:
  – Word-aligned sentence pair
  – Constituency parse trees on one or both sides

• Outputs:
  – Set of SCFG rules derivable from the inputs, possibly according to some constraints

• Implemented by:
  Hiero [Chiang 2005]  GHKM [Galley et al. 2004]
  Chiang [2010]       Stat-XFER [Lavie et al. 2008]
  SAMT [Zollmann and Venugopal 2006]
SCFG Grammar Extraction

• Our goals:
  – Support for two parse trees by default
  – Extract greatest number of syntactic rules...
  – Without violating constituent boundaries

• Achieved with:
  – Multiple node alignments
  – Virtual nodes
  – Multiple right-hand-side decompositions

First grammar extractor to do all three
Ma mère avait toujours aimé les voitures bleues.
Basic Node Alignment

- Word alignment consistency constraint from phrase-based SMT
Basic Node Alignment

- Word alignment consistency constraint from phrase-based SMT
Virtual Nodes

- Consistently aligned consecutive children of the same parent
Virtual Nodes

- Consistently aligned consecutive children of the same parent
- New intermediate node inserted in tree
Virtual Nodes

- Consistently aligned consecutive children of the same parent
- New intermediate node inserted in tree
- Virtual nodes may overlap
- Virtual nodes may align to any type of node
Syntax Constraints

- Consistent word alignments ≠ node alignment
- Virtual nodes may not cross constituent boundaries
Multiple Alignment

- Nodes with multiple consistent alignments keep all of them
Basic Grammar Extraction

• Aligned node pair is LHS; aligned subnodes are RHS

\[ NP::NP \rightarrow [\les \ N^{1} \ A^{2}]::[JJ^{2} \ NNS^{1}] \]
\[ N::NNS \rightarrow [\text{voitures}]::[\text{cars}] \]
\[ A::JJ \rightarrow [\text{bleues}]::[\text{blue}] \]
**Multiple Decompositions**

- All possible right-hand sides are extracted

\[
\begin{align*}
\text{NP} & \rightarrow [\text{les } \text{N}^1 \text{ A}^2]::[\text{JJ}^2 \text{ NNS}^1] \\
\text{NP} & \rightarrow [\text{les } \text{N}^1 \text{ bleues}]::[\text{blue NNS}^1] \\
\text{NP} & \rightarrow [\text{les voitures } \text{A}^2]::[\text{JJ}^2 \text{ cars}] \\
\text{NP} & \rightarrow [\text{les voitures bleues}]::[\text{blue cars}] \\
\text{N} & \rightarrow [\text{voitures}]::[\text{cars}] \\
\text{A} & \rightarrow [\text{bleues}]::[\text{blue}] \\
\end{align*}
\]
Multiple Decompositions

NP::NP → [les N+AP1]:[NP1]
NP::NP → [D+N1 AP2]:[JJ2 NNS1]
NP::NP → [D+N1 A2]:[JJ2 NNS1]
NP::NP → [les N1 AP2]:[JJ2 NNS1]
NP::NP → [les N1 A2]:[JJ2 NNS1]
NP::NP → [D+N1 bleues]:[blue NNS1]
NP::NP → [les N1 bleues]:[blue NNS1]
NP::NP → [les voitures AP2]:[JJ2 cars]
NP::NP → [les voitures A2]:[JJ2 cars]
NP::NP → [les voitures bleues]:[blue cars]
D+N::NNS → [les N1]:[NNS1]
D+N::NNS → [les voitures]:[cars]
N+AP::NP → [N1 AP2]:[JJ2 NNS1]
N+AP::NP → [N1 A2]:[JJ2 NNS1]
N+AP::NP → [N1 bleues]:[blue NNS1]
N+AP::NP → [voitures AP2]:[JJ2 cars]
N+AP::NP → [voitures A2]:[JJ2 cars]
N+AP::NP → [voitures bleues]:[blue cars]
N::NNS → [voitures]:[cars]
AP::JJ → [A1]:[JJ1]
AP::JJ → [bleues]:[blue]
A::JJ → [bleues]:[blue]
Constraints

• Max rank of phrase pair rules
• Max rank of hierarchical rules
• Max number of siblings in a virtual node
• Whether to allow unary chain rules

\[
\text{NP::NP} \rightarrow [\text{PRO}^1]::[\text{PRP}^1]
\]

• Whether to allow "triangle" rules

\[
\text{AP::JJ} \rightarrow [\text{A}^1]::[\text{JJ}^1]
\]
# Comparison to Related Work

|                | Tree Constr. | Multiple Aligns | Virtual Nodes | Multiple Decomp. |
|----------------|--------------|-----------------|---------------|------------------|
| Hiero          | No           | —               | —             | Yes              |
| Stat-XFER      | Yes          | No              | Some          | No               |
| GHKM           | Yes          | No              | No            | Yes              |
| SAMT           | No           | No              | Yes           | Yes              |
| Chiang [2010]  | No           | No              | Yes           | Yes              |
| This work      | Yes          | Yes             | Yes           | Yes              |
Experimental Setup

- Train: FBIS Chinese–English corpus
- Tune: NIST MT 2006
- Test: NIST MT 2003
Extraction Configurations

• Baseline:
  – Stat-XFER exact tree-to-tree extractor
  – Single decomposition with minimal rules

• Multi:
  – Add multiple alignments and decompositions

• Virt short:
  – Add virtual nodes; max rule length 5

• Virt long:
  – Max rule length 7
## Number of Rules Extracted

|        | Tokens |                  | Types |                  |
|--------|--------|------------------|-------|------------------|
|        |        | Phrase | Hierarc. | Phrase | Hierarc. |
| Baseline | 6,646,791 | 1,876,384 | 1,929,641 | 767,573 |
| Multi   | 8,709,589 | 6,657,590 | 2,016,227 | 3,590,184 |
| Virt short | 10,190,487 | 14,190,066 | 2,877,650 | 8,313,690 |
| Virt long | 10,288,731 | 22,479,863 | 2,970,403 | 15,750,695 |
Multiple alignments and decompositions:

- Four times as many hierarchical rules
- Small increase in number of phrase pairs
### Number of Rules Extracted

|                  | Tokens          |                  | Types           |                  |
|------------------|-----------------|-----------------|-----------------|-----------------|
|                  | Phrase | Hierarc. | Phrase | Hierarc. |
| Baseline         | 6,646,791 | 1,876,384 | 1,929,641 | 767,573 |
| Multi            | 8,709,589 | 6,657,590 | 2,016,227 | 3,590,184 |
| Virt short       | 10,190,487 | 14,190,066 | 2,877,650 | 8,313,690 |
| Virt long        | 10,288,731 | 22,479,863 | 2,970,403 | 15,750,695 |

- Multiple decomp and virtual nodes:
  - 20 times as many hierarchical rules
  - Stronger effect on phrase pairs
  - 46% of rule types use virtual nodes
### Number of Rules Extracted

|                  | Tokens        |                      | Types         |                      |
|------------------|---------------|----------------------|---------------|----------------------|
|                  | Phrase        | Hierarc.             | Phrase        | Hierarc.             |
| Baseline         | 6,646,791     | 1,876,384            | 1,929,641     | 767,573              |
| Multi            | 8,709,589     | 6,657,590            | 2,016,227     | 3,590,184            |
| Virt short       | 10,190,487    | 14,190,066           | 2,877,650     | 8,313,690            |
| Virt long        | 10,288,731    | 22,479,863           | 2,970,403     | 15,750,695           |

- Proportion of singletons mostly unchanged
- Average hierarchical rule count drops
Rule Filtering for Decoding

• All phrase pair rules that match test set
• Most frequent hierarchical rules:
  – Top 10,000 of all types
  – Top 100,000 of all types
  – Top 5,000 fully abstract + top 100,000 partially lexicalized

\[
\text{VP::ADJP} \rightarrow [VV^1 \ VV^2]::[RB^1 \ VBN^2]
\]
\[
\text{NP::NP} \rightarrow [2000 \ 年 \ NN^1]::[\text{the} \ 2000 \ NN^1]
\]
## Results: Metric Scores

- **NIST MT 2003 test set**

| System       | Filter | BLEU  | METR  | TER   |
|--------------|--------|-------|-------|-------|
| Baseline     | 10k    | 24.39 | 54.35 | 68.01 |
| Multi        | 10k    | 24.28 | 53.58 | 65.30 |
| Virt short   | 10k    | 25.16 | 54.33 | 66.25 |
| Virt long    | 10k    | 25.74 | 54.55 | 65.52 |

- **Strict grammar filtering:** extra phrase pairs help improve scores
### Results: Metric Scores

- **NIST MT 2003 test set**

| System      | Filter    | BLEU   | METR   | TER    |
|-------------|-----------|--------|--------|--------|
| Baseline    | 5k+100k   | 25.95  | 54.77  | 66.27  |
| Virt short  | 5k+100k   | 26.08  | 54.58  | 64.32  |
| Virt long   | 5k+100k   | 25.83  | 54.35  | 64.55  |

- **Larger grammars: score difference erased**
Conclusions

• Very large linguistically motivated rule sets
  – No violating constituent bounds (Stat-XFER)
  – Multiple node alignments
  – Multiple decompositions (Hiero, GHKM)
  – Virtual nodes (< SAMT)

• More phrase pairs help improve scores

• Grammar filtering also matters
Future Work

- Filtering to limit derivational ambiguity
- Filtering based on content of virtual nodes
- Reducing the size of the label set
  - Original: 1,577
  - With virtual nodes: 73,000
References

- Chiang (2005), “A hierarchical phrase-based model for statistical machine translation,” ACL
- Chiang (2010), “Learning to translate with source and target syntax,” ACL
- Galley, Hopkins, Knight, and Marcu (2004), “What’s in a translation rule?,” NAACL
- Lavie, Parlikar, and Ambati (2008), “Syntax-driven learning of sub-sentential translation equivalents and translation rules from parsed parallel corpora,” SSST-2
- Zollmann and Venugopal (2006), “Syntax augmented machine translation via chart parsing,” WMT