CCG Contextual Labels in Hierarchical Phrase-Based SMT

Hala Almaghout, Jie Jiang, Andy Way

CNGL, School of Computing, Dublin City University
Outline

- Introduction
- Related Work
- Our Approach
- Experiments
- Analysis
- Conclusion
- Future Work
Hierarchical Phrase-Based SMT (Chiang, 2005)

- **Parallel Corpus**
  - Source Text
  - Target Text
  - Training
  - Tuning
  - Development set

- **Synchronous Context-Free Grammar**
  - \[ \sum \lambda \log(p_i) \]

- **He bought a ticket from Ankara to Dublin**
  - a ticket
  - Ankara
  - Dublin
  - from Ankara to Dublin
  - He bought a ticket
  - He bought a ticket from Ankara to Dublin
He bought a ticket from Ankara to Dublin

Hierarchical Rule Extraction

S \rightarrow ( X2 \rightarrow ( X1, He bought a ticket from X1 to X2 ) )

source side nonterminals

source side nonterminals

Left hand side nonterminal

target side

target side nonterminals

He bought a ticket from Ankara to Dublin

from Ankara to Dublin

a ticket from from Ankara to Dublin

He bought a ticket from Ankara to Dublin
There are no syntactic constraints imposed on phrases replacing nonterminals in hierarchical rules.

Solution:

Syntax Augmented Machine Translation (SAMT) (Zollmann and Venugopal, 2006)
Syntax Augmented Machine Translation (SAMT)

He bought a ticket from Ankara to Dublin

S → (NP

VP\VBD → (NP

PP+PP → (NP

VP\VBD → (PP+PP

S → VP\VBD

X → (Ankara to Dublin)

S → (He bought a ticket from X)
He bought a ticket from Ankara to Dublin.

He bought a ticket from from Ankara to Dublin.

He bought a ticket from Ankara to Dublin.

He bought a ticket from Ankara to Dublin.
He bought a ticket from Ankara to Dublin.

SAMT Rule Extraction

S → (NP を MEN ヒブから POV 向けた BO を NNP を MEN 向けた BO から NNP を MEN 向けた － , a ticket Men MEN 向けた BO から NNP を MEN 向けた BO から NNP を MEN 向けた － , from NNP 向けた BO から NNP を MEN 向けた － , a ticket PP+PP )

VP/VBD → (NP を MEN 向けた BO から NNP を MEN 向けた BO から NNP を MEN 向けた － , a ticket PP+PP )

PP+PP → (NP を MEN 向けた BO から NNP を MEN 向けた BO から NNP を MEN 向けた － , from NNP 向けた BO から NNP を MEN 向けた － , a ticket PP+PP )

S → ( VP/VBD を MEN 向けた BO から NNP 向けた BO から NNP 向けた － , a ticket PP+PP )

X → (NP を MEN 向けた BO から NNP 向けた BO から NNP 向けた － , from NNP 向けた BO から NNP 向けた － , a ticket PP+PP )

S → (X を MEN 向けた BO から NNP 向けた BO から NNP 向けた － , a ticket PP+PP )

He bought a ticket from X.
SAMT Limitations

- **Label sparsity**: using many different nonterminal labels. This results from using SAMT combinatory operators.

  \[ S \rightarrow (X, \text{He bought a ticket}) \]

  \[ S \rightarrow (PP, \text{He bought a ticket}) \]

  \[ S \rightarrow (PP+PP, \text{He bought a ticket}) \]

  \[ S \rightarrow (SBAR, \text{He bought a ticket}) \]

- **Label coverage**: failing to find a syntactic label expressing the syntactic function of some of SMT phrases.

- **Label accuracy**: how accurate the SAMT labels are in reflecting the real syntactic function of the phrases.

  - SAMT has larger translation model.
  - Low-probability rules weaken the system’s ability to generalize, and damage the performance of the system.
Producing more grammatical translation by imposing syntactic constraints on nonterminal replacement

Constituency grammar rigid structures $\rightarrow$ Label coverage
SAMT combinatory operators $\rightarrow$ Label sparsity & accuracy

Using Combinatory Categorial Grammar (CCG) to label target-side phrases instead of constituency grammar
+ CCG more flexible and richer structures $\rightarrow$ Label coverage
+ CCG supertags reflect rich syntactic information at the lexical level $\rightarrow$ Label sparsity & accuracy
+ CCG is efficiently parsed
Combinatory Categorial Grammar (CCG)

CCG = Lexicon + Combinatory Operators

Lexicon:
- He → NP
- Walks → S\NP
- reads → (S\NP)/NP, S\NP
- a → NP/N
- book → N, (S\NP)/NP
- ...

Combinatory Operators:
- FA: X/Y → Y
- BA: Y X\Y → Y
- FC: X/Y Y/Z → X/Z
- BC: Y\Z X\Y → X\Z

He walks NP S\NP

He reads a book NP (S\NP)/NP NP/N N

S\NP

S

Primitive Category

Complex Category of the form X\Y
CCG-Augmented Hierarchical Phrase-Based SMT

Coarse-grained Chomsky Normal Form (CCG)

Syntactic Analysis & Annotation

Parallel Corpus

Source Text

Target Text

Training

Development set

Translation model

Translation

Output

He bought a ticket from Ankara to Dublin

S[dcl] → NP → NP

He bought a ticket from NP to NP

NP → (NP → NP)

a ticket from NP to NP

NP

Ankara

NP

bought

(S[dcl])/NP

to (NP/NP)/NP

ticket N

from ((NP/NP)/(NP/NP))/NP

Dublin NP

Σ ∑ log(πi)

 Arabic Text

دن إاشتراي بطاقة من اتراكا إلى دبلن

He bought a ticket from Ankara to Dublin
He bought a ticket from Ankara to Dublin
CCG-Augmented HPB SMT vs. SAMT

CCG-Augmented HPB system outperformed SAMT in terms of BLEU score on Arabic—English and Chinese—English news translation.

- CCG supertag labels are less sparse and are able to label more phrases than SAMT labels.
- CCG-Augmented HPB system could not outperform the HPB baseline.
- CCG supertag labels still suffer from label sparsity problem.

Number of different syntactic labels vs. Percentage of phrase-pairs without syntactic label.
Solution

- **Softening** CCG supertags labels by employing part of the information represented in them.
- Two softening methods:
  - CCG contextual labels
  - Feature-removed CCG labels
- Goal:
  - Reduce label sparsity.
  - Loosen syntactic constraints.
- However, this comes at the expense of the accuracy of the syntactic labels.
He bought a ticket from Ankara to Dublin.
CCG contextual labels can be extracted directly from CCG supertags without the need to parse the phrase.

He bought a ticket from Ankara to Dublin.
Feature-removed CCG Labels

- S[dcl]\NP → Declarative Verb Phrase
- S[to]\NP → Infinitival Verb Phrase
- S[pt]\NP → Past Participle Verb Phrase

S\NP
Experiments

- **Language pairs**: Arabic—English and Chinese—English
- **Data used**: from the news and travelling speech expressions domains (IWSLT 2010 evaluation campaign).

<table>
<thead>
<tr>
<th></th>
<th>News</th>
<th>IWSLT</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE (ar-en)</td>
<td>48065</td>
<td>21484</td>
</tr>
<tr>
<td>CE (zh-en)</td>
<td>51044</td>
<td>63234</td>
</tr>
</tbody>
</table>

- **Baseline Systems**:
  - The PB baseline system: built using the Moses PB Decoder.
  - The HPB baseline system: built using the Moses Chart Decoder.

- **CCG-based Systems**:
  - CCG Context: uses CCG contextual labels.
  - CCG: uses CCG supertag labels.
  - CCG (s): uses feature-removed CCG supertag labels.
  - CCG Context (s): uses feature-removed CCG contextual labels.
BLEU Scores for Arabic—English Experiments

**Ar-En News**

- CCG: 24
- PB: 23
- CCG Context: 23
- HPB: 21
- CCG (s): 20

**Ar-En IWSLT**

- HPB: 54
- CCG Context (s): 53
- CCG Context: 53
- CCG (s): 52
- CCG: 51
- PB: 50

- **0.56 BLEU points Statistically Significant at p-level=0.05**
- **0.06 BLEU points Only 20k**
BLEU Scores for Chinese—English Experiments

**Zh-En News**

- CCG (s): 24
- PB: 24
- HPB: 24
- CCG Context: 24
- CCG: 24
- CCG Context (s): 24

**Zh-En IWSLT**

- CCG Context: 51
- HPB: 51
- CCG Context (s): 51
- CCG (s): 49
- CCG: 47
- PB: 45

0.03 BLEU points

0.56 BLEU points
Label Sparsity

simplified CCG labels are less sparse than CCG supertag labels
Conclusions

- CCG label simplification demonstrated to be promising
  - At least one of the systems which use simplified CCG labels achieved better BLEU score than the CCG supertags HPB baseline.
  - Simplified CCG label systems were the best performing systems on all but AE IWSLT experiment.

- In comparison with CCG supertag labels, CCG contextual labels demonstrated to be:
  - less sparse
  - easier to extract than CCG supertags

- Simplification schemes did not show consistent improvement over baseline systems on a specific language pair or corpus type.
Future Work

- Conducting a thorough evaluation of *CCG label simplification schemes* using *larger training corpora* and on *more language pairs*.
- Examining the effect of *source language segmentation* on the performance of CCG-based systems.
- Using *system combination* on CCG-based systems to obtain a better performing system.
- Conducting a *manual analysis* on selected sentences to examine the effect of using CCG-based labels on producing more grammatical translations.
Thanks for your Attention!
Questions??