Maximum Entropy Translation Model in Dependency-Based MT Framework

David Mareček, Martin Popel, Zdeněk Žabokrtský

Charles University in Prague
Institute of Formal and Applied Linguistics

PIRE meeting
July 15, 2010, Uppsala, Sweeden
Outline

- TectoMT system introduction
- Translation dictionaries
- Experiments and results
TectoMT

- Analysis-transfer-synthesis translation system
  - Transfer on the level of deep-syntax (tectogrammatical trees)
Tectogrammatical tree

- Dependency tree, where only content words have their own nodes
- Other words (function words) are expressed within the respective content nodes in the form of their attributes
  - Function words: articles, prepositions, auxiliary verbs, modal verbs, punctuation marks, ...
- Three main attributes of the nodes we need:
  - T-lemma
  - Formeme - surface morphosyntactic form of the node
  - Grammatemes - morphological categories
She has never been to Sydney.
Tectogrammatical attributes

- **T-lemma**
  - mother, read, #PersPron, take_off, put_on, look_after

- **Formeme** – surface morphosyntactic form of the node
  - English: n:subj, n:obj, n:of+X, n:x+ago, adj:attr, v:fin, adv:, ...
  - Czech: n:1, n:2, n:na+4, v:inf, adj:attr, ...

- **Grammatemes** – necessary morphological categories
  - gender, number, person, tense, verb modality, degree of comparison, ...
Why is tectogrammatics good for transfer

- Tectogrammatical trees of corresponding Czech and English sentences are much more similar than their surface shapes
  - Contain content words only
  - Contain also entities elided on surface
We can assume that the target structure will be the same (isomorphic) as the source structure than we can simply translate each node in 1:1 manner. Of course, there exists 1:2, 2:1 and other mappings:

- ice cream = zmrzlina
- mother in law = tchýně

There are not much of them. These mappings can be solved separately by a special dictionary.

When evaluated, only 8% of errors were caused by this assumption of isomorphism.
The transfer process

- The topology of the tree and grammemes (such as person, number, gender, tense, etc.) are preserved.
- For each node, its t-lemma and formeme are translated separately.
  - For each t-lemma/formeme the set of translation variants is generated from dictionaries.
The transfer process

- Each translation variant has a probability assigned from dictionaries
- The translation variants are pruned
The transfer process

- The optimal combination of lemmas and formemes is chosen using TreeLM
  - Hidden-Tree-Markov-Models (HMTM)
  - Viterbi search

```
mother
n:attr
tongue
n:obj
matka
n:attr
materšký
rodný
n:2
adj:attr
n:attr
jazyk
spice
mluva
n:4
n:1
n:o+6
```

```
0.32
0.15
0.12
0.61
0.17
0.02
0.32
0.15
0.12
0.21
0.29
0.08
```
Outline

- TectoMT system introduction
- Translation dictionaries
- Experiments and results
Translation dictionaries

- For a given source lemma/formeme it returns a set of translation hypotheses (with probabilities)

- Translation of lemmas:
  - Static dictionary
  - Context (MaxEnt) dictionary
  - Derivative dictionary (rule-based)

- Translation of formemes:
  - Static dictionary
  - Context (MaxEnt) dictionary
Static dictionary

- Simple dictionary extracted from aligned parallel treebank
  - Extracted all aligned pairs of Czech and English nodes
  - Maximum likelihood estimation
- \( p(c|e) = \frac{\text{count}(c,e)}{\text{count}(e)} \)
- Used both for transliteration of lemmas and formemes
Context (MaxEnt) dictionary

- Uses also the context of the source node and other attributes

\[ p(y|x) = \frac{1}{Z(x)} \exp \sum_{i} \lambda_i f_i(x, y) \]

- One MaxEnt model is trained for each source lemma
- Source context features used (x):
  - Local tree context
  - Local linear context
  - Morphological and syntactic categories
Examples of features:

- Tense of governing node = past
- Lemma of the previous node = „cut“
- Child node formeme = „n:by+X“
- Has left child = 1
The grass in front of our house has been cut.

source_lemma="cut" & child_lemma="grass" & tense="ant"
→ target_lemma="posekat"
Derivative dictionaries

- Translation of unknown words using their derivation
  - Translation of adverbs through adjectives:
    - interestingly → interesting → zajímavý → zajímavě
  - Translation of adjectives through verbs
    - translatable → translate → přeložit → přeložitelný
Derivative dictionaries

- Translate prefixes separately
  - using lexicon of prefixes
    - Multi-core → vícejádrový
    - Neoclassicism → neoklasicismus

- If we recognize suffix in an unknown word, we translate the suffix only
  - Using lexicon of suffixes
    - Geocentrism → geocentrumus
- Rules of translation some of hyphen-compounds
  - First part is a number the second is a noun, which is translated as an adjective
    - Two-litre → dvoulitrový
    - 45-year-old → pětačtyřicetiletý
    - Three-fifths → třípětinový

- What probability to return?
  - 1, because it is the only variant we have for the unknown source lemma
The described dictionaries are combined in the following way:

- MaxEnt
- Human
- Static

Linear interpolation

Backoff

- Hyphen compounds
- Deadjectival adverbs
- Deverbal adjectives
- Prefixes
- Suffixes
Outline

- TectoMT system introduction
- Translation dictionaries
- Experiments and results
Experiments and Results

- Resources:
  - For TM: Czech-English parallel corpus CzEng 0.9, approx. 60 megawords on both sides, analyzed up to tectogrammatical layer and aligned
  - For LM: Czech National Corpus, 800 megawords
  - Evaluation data from WMT 2010 test set (2489 sentences)

<table>
<thead>
<tr>
<th>Dictionary used</th>
<th>BLEU</th>
<th>NIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static only (MLE)</td>
<td>11.67</td>
<td>5.023</td>
</tr>
<tr>
<td>Static + MaxEnt</td>
<td>12.48</td>
<td>5.234</td>
</tr>
<tr>
<td>Static + MaxEnt + Derivative</td>
<td>12.58</td>
<td>5.250</td>
</tr>
</tbody>
</table>
Conclusions

- TectoMT – the analysis-transfer-synthesis MT system with transfer over the deep-syntax was described
- We have focused on the system of translation dictionaries:
  - Static dictionary (MLE)
  - Context dictionary (Maximum Entropy)
  - Derivational dictionary (rule-based)
- We have shown that all the dictionaries improved the quality of machine translation
  - almost 1 BLEU point improvement
Thank you for your attention!