Generation or transfer: this is still the question

Nianwen Xue
Brandeis University
(Joint work with Dun Deng)
2014-07-21
PIRE Workshop
Outline

• The trend to go deeper: dependency tree to Abstract Meaning Representation (AMR)
• A shallower alternative for MT: a hierarchically aligned Chinese-English parallel treebank
Dependency representation revolves around the notion of “head”

- Root is ‘is’
- ‘is’ is the head of ‘Ivan’
- ‘is’ is the head of ‘dancer’
- ‘dancer’ is the head of ‘best’
- ‘dancer’ is the head of ‘the’
Dependency representation revolves around the notion of “head”

Root is “office”
“office” is the head of “of”
“of” is the head of “Chair”
“Chair” is the head of “the”
It’s not always this clean

Prague dependency

Moscow dependency
Universal Stanford dependency (USD)

Relation between content words mediated by function words

Direct links between content words
Universal Stanford dependency (USD)

Relation between content words mediated by function words

Direct links between content words
Universal Stanford dependency (USD)

Relation between content words mediated by function words

Direct links between content words
Abstract Meaning Representation (AMR)

Function words kept

Function words thrown away
(or mapped to abstract concepts or relations, or attributes)
Abstract Meaning Representation (AMR)

Function words kept

Function words thrown away
(or mapped to abstract concepts, attributes, or relations)
Abstract Meaning Representation (AMR)

Ivan is the best dancer

Function words kept
Lexical Integrity respected

Function words thrown away
(or mapped to abstract concepts, attributes, or relations)
Lexical integrity violated
The trend

• From “standard” dependency to Stanford dependency to AMR, the community is moving towards increasingly “deeper”, more abstract representations
The case for AMR as a representation for MT
The case for AMR as a representation for MT
当你需要他时, 霍默辛普森在哪里？
Where is Homer Simpson when you need him?
这是一个大叫“噢哦！”的时刻。
This is a major `` D'oh! " moment .
The representation calculus

• Deeper, more abstract representations lead to better alignment

• But they come at a cost for MT purposes:
  – Deeper analyzers tend to be less accurate
  – For MT, AMR necessitates a generation step: you have to get back the function words you’ve thrown away or mapped to abstract concepts, attributes, or relations
    • Source string ➔ (source AMR) ➔ target AMR ➔ target string

• Less of a problem for applications such as IE, QA, etc. where the generation step is less involved
Hierarchical alignment between constituent parse trees

Let’s not throw away the function words. Just don’t align them (at the word level).

Assume neural networks do not solve all the problems in MT and representation is still relevant.


Word alignment

provide you with an explanation

给 你 提供 解释
Existing word alignment practice

- Function words are attached to a “head”
  - Determiners (e.g., “the”, “a”) and Chinese classifiers (e.g., 个) are attached to a noun before they are aligned
  - Auxiliary verbs, tense, aspect markers are attached to their main verbs
  - Prepositions are attached to their NP object
  - Complementizers attached to matrix verbs,
  - Relative pronouns attached to head of relative clauses

- Function words (or punctuation) that anchor larger patterns are aligned at the word level as well
  - Chinese comma aligned with “and”
  - Chinese sentence-final markers aligned with punctuation marks
  - Chinese 的, a modifier marker, attached to and aligned with various things

- Pronouns are linked to their coreferent before they are aligned
Linguistic problems

• Cram too much into word-level alignment
Computational problems

• Creates pervasive *one-to-many* or *many-to-many* alignments, where the “many” are often non-consecutive, making the automatic alignment difficult.

• Severs crucial dependencies and creates many haphazard alignments, increases ambiguity
Attaching function words to their “heads”

你 -- you
给你 -- you

解释 – an explanation
解释 – with an explanation
Attaching function words to their “heads”

提供给你解释
provide sb with sth

解释 – an explanation
解释 – with an explanation
Spurious ambiguity magnified

Eat apples <> 吃 苹果
Eat an apple <> 吃 苹果
Eat the apple <> 吃 苹果
Fond of apples <> 喜欢 苹果
Talk about apples <> 谈论 苹果
Provide them with apples <> 给 他们 提供 苹果
Hierarchical alignment

• Aligning function words in their syntactic context
  – Leaving function words ("glue words") that signal syntactic patterns (的, "so that") or semantic attributes ("a", "the") unaligned at the word level, and migrating their alignment to the phrase level

• What is the proper syntactic context?
Word alignment

提供解释

provide you with an explanation
Hierarchical alignment

provide you with an explanation

給你提供解釋
Hierarchical alignment constraints

• If Node $n_c$ is aligned to Node $n_e$, then the descendants of $n_c$ can only be aligned to the descendants of $n_e$
• If Node $n_c$ is aligned to Node $n_e$, then the ancestors of $n_c$ can only be aligned to the ancestors of $n_e$
• A non-terminal node can only be aligned once

(Tinsley et al, 2007)
Hierarchical alignment

provide you with an explanation

Hierarchical alignment
Hierarchical alignment

provide you with an explanation

给 你 提供 解释

provide NP with VP NP

provide VP PP NP

provide VP P NP

provide VP P IN NP

provide VP VB NP

provide VP VB NP

provide VP VV VP

provide VP VV VP
Guiding principles for determining the alignment of non-terminal nodes

• Provide enough context for the alignment
  – Alignment should not sever crucial dependencies or patterns
  – Reduces the number of haphazard alignments to its minimum

• Minimum alignment
  – Aligned (non-terminal) nodes dominate only crucial dependencies, nothing more
Hierarchical alignment

provide you with an explanation

give you an explanation
Hierarchical alignment

provide you with an explanation

[你 <-> you]

[解释 <-> an explanation]

[给X₁提供X₂ <-> provide X₁ with X₂]
Alignment procedure

- Take a portion of the Penn Chinese TreeBank (CTB) that has English translations that have also been manually parsed (PTB) style and word aligned.
- Revising the word alignments, and then align the non-terminal nodes, using a web-based tool developed by IBM.
Attaching a determiner to its head

S3878
Hierarchically aligned

<table>
<thead>
<tr>
<th>TOP</th>
<th>IP</th>
<th>NP_SBJ</th>
<th>PN</th>
<th>VP</th>
<th>PP_ADV</th>
<th>P</th>
<th>LNP</th>
<th>QP</th>
<th>OD</th>
<th>LCLP</th>
<th>M</th>
<th>LN</th>
<th>NN</th>
<th>LNP</th>
<th>NN</th>
<th>LNP</th>
<th>AD</th>
<th>VP</th>
<th>VA</th>
<th>PU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>g</td>
<td>一切</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>extra</td>
<td>和</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>g</td>
<td>第一</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>extra</td>
<td>个</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>extra</td>
<td>路障处</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>g</td>
<td>看上去</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>extra</td>
<td>一样</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Everything: looked 3 IN PP CLP first 4 DT NP UFW roadblock 6 NN .

S3878
Attaching 把 to its NP “object”
Hierarchically aligned
Verb prepositions/particles

S3877

1. g 他
2. g 也
3. g 没
4. g 能
5. g 拿出
6. extra 会费
7. g 来
8. g 。

he  g  1
had  g  2
n't  g  3
been g  4
able g  5
to extra  6
come g  7
up g  8
with g  9
the g  10
dues g  11
,  extra  12
either g  13
.  g  14
Hierarchically aligned
Auxiliary verbs
Hierarchically aligned

<table>
<thead>
<tr>
<th>TOP-CP</th>
<th>IP-NP_SBJ</th>
<th>PN</th>
<th>1</th>
<th>那</th>
<th>2</th>
<th>extra</th>
<th>是</th>
<th>3</th>
<th>g</th>
<th>怎么</th>
<th>4</th>
<th>g</th>
<th>发生</th>
<th>5</th>
<th>extra</th>
<th>的</th>
<th>6</th>
<th>g</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVP</td>
<td>VC</td>
<td></td>
<td>2</td>
<td>extra</td>
<td>3</td>
<td>g</td>
<td>怎么</td>
<td>4</td>
<td>g</td>
<td>发生</td>
<td>5</td>
<td>extra</td>
<td>的</td>
<td>6</td>
<td>g</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVP-ADVP_WH-AD</td>
<td>g</td>
<td>happen</td>
<td>g</td>
<td>4VB</td>
<td>VP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| how    | g | 1 | WRB-WHADVP | 1 | -SBARQ | TOP |
| did    | extra | 2 | VBD | SC |
| that   | g | 3 | DT | NP_SBJ |
| happen | g | 4 | VB | VP |
Are the PTB and CTB trees compatible?

• Not always. Some of the incompatibilities are well-documented in MT literature

• Three types of incompatibilities:
  – Incompatibilities between lexico-semantic differences between the two languages
  – Incompatibilities caused by translation-related reasons
    • 5 cases out of 500 sentence pairs
  – Incompatibilities caused by bracketing annotation
    • 20 cases out of 500 sentence pairs

• The last one we can do something about
Figure 2: Translation of idiomatic expressions.
Divergent annotation standards

(a) Penn English TreeBank structure

(b) Chinese TreeBank structure

Figure 4: Differences in treebank annotation standards. XP = \{NP, PP, ADVP, S\}
Divergent annotation standards

The Senate reopened the budget talks last week.
Can this be done quickly and consistently?

• Annotators with no prior linguistic training can perform this kind of alignment with good consistency given proper training and guidelines
  – 87% Inter-annotator agreement for (non-terminal) node alignment

• Tree synchronization requires linguistic training
  – We currently just mark places of tree divergence

• Status:
  – Double annotation of about 10,000 sentence pairs
Use cases for hierarchical alignment

• Extracting Hierarchical rules
  – From an unaligned function word, find its alignment domain by identifying its closest ancestor node N that is aligned.
  – From N, all descendant nodes that are not node-aligned themselves are visible as potential triggers of a hierarchical rule.

• Alignment-driven flattening of syntactic trees
  – We can assume the unaligned nodes are unimportant to MT and can be automatically eliminated to create a simplified tree for MT purposes.
[可 禁止 $X_1$ $X_2$ $\iff$ $X_1$ is prohibited from $X_2]$
把 construction

[把 X 定为非法因素 <-> outlaw X]
[将 $X_1$，这样 $X_2 \leftrightarrow \text{will } X_1 \text{ so that } X_2$]
Conjunction

[X , Y <> X and Y]
[ $X_1$ 为什么 不 $X_2$ <-> Why do n’t $X_1 X_2$]
那X为什么Y呢？ <> Then why do X Y?
[认为 我们 不必 X 都 看到 Y <> do n’t think we need see Y X]
Summary

• Strategically select non-terminal node pairs to align
  • Effectively creates an hierarchical partition for each sentence

• Harmonize word alignment and (non-terminal) node alignment
  – Content words are (generally) aligned the word level
  – Function words are (generally) aligned via the alignment of non-terminal nodes

• Synchronize parallel parse trees
  – Address divergences in PTB and CTB annotation styles
Cases for and against hierarchical alignment for MT

• Supports transfer-based approach to MT, no need for a separate generation step. More steps in the pipeline mean more chances for errors.
• Can be readily exploited by existing statistical techniques (supervised synchronized parsing?)
• But need to be done for each language pair. May not work for some language pairs.
Conclusions

- Familiar tradeoffs between depth of representation (therefore better alignment) and processing difficulty (analysis, generation)
- A hierarchically aligned parallel corpus might be a viable alternative as a representation for MT
Acknowledgement

• Libin Shen for inspirations
• Tool support by the IBM Multilingual NLP Technologies Group: Niyu Ge, Abe Ittycheriah, Salim Roukos
• Funded by DARPA BOLT Program via IBM