Textual entailment inference in machine translation

Ido Dagan and Shachar Mirkin

Workshop on Machine Translation and Morphologically-rich Languages

Haifa, January 2011

Joint work with: Wilker Aziz, Nicola Cancedda, Marc Dymetman, Lucia Specia and Idan Szpektor
• Textual Entailment

• Unified view of entailment and MT

• Handling OOV in MT with entailment
  • ACL 09 - entailment contribution to MT
  • EAMT10 - Integration into standard SMT workflow
The Tunisian embassy in Switzerland was attacked. Fire bombs were thrown at the Tunisian embassy in Bern.

**Core task:** Can the meaning of a target textual assertion (hypothesis, $H$) be inferred from a given text ($T$)?

- In this case: $T$ entails $H$ ($T \Rightarrow H$)

Paraphrasing is bi-directional entailment.

- $T'$ and $T$ mutually entail each other (paraphrases) ($T \Leftrightarrow T'$)
Reducing applications’ inferences to entailment

- **Question Answering**

<table>
<thead>
<tr>
<th>Question</th>
<th>Expected answer template</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who founded Wikileaks?</td>
<td>X founded Wikileaks</td>
</tr>
</tbody>
</table>

- **Similar setting: Information Extraction**
  - X founded Y
The task

Given student’s textual answer to a system’s question – assess the answer relative to a reference answer.

Question: An object has to move to produce sound. Do you agree?

Reference: Agree. Vibrations are movements and vibrations produce sound.

Student’s answer: Yes because it has to vibrate to make sounds.

The entailment perspective:

Student answer should paraphrase or entail the reference.

A similar setting to MT evaluation (more soon)
The RTE Challenges

• Based on application scenarios and data

• Annually, since 2005 (RTE-7 expected in 2011)

• Very successful challenges, world wide:
  • Dozens of participating groups so far (~20 each year)
  • Hundreds of downloads

• Since RTE-4 (2008) – under NIST
  • New Text Analysis Conference (TAC, sister for TREC)

• Current trend:
  • Reflect RTE potential utility for other TAC applications
    • Update summarization task, KBP slot filling
Textual entailment ≈ human reading comprehension

• From an English matriculation exam (Israel, 2010):

Norman: Why do you think teenagers smoked less in 2003?

Dr. Clark: Anti-smoking advertisements convinced teenagers not to start smoking. Teenagers also got more information about the dangers of smoking from parents, teachers and friends. In addition, in those years, people were not allowed to smoke in public places any more.
Entailment and MT
Entailment for MT evaluation (apropos of tutoring)

• Kauchak & Barzilay, 2006
  • **Assumption:** The translation can be a paraphrase of the reference rather than an exact match
  • **Method**
    • Paraphrasing the translation such that it becomes more similar to the reference

• Potential extension to directional entailment:
  • If the *reference directionally-entails the system translation*, the translation may still be useful, though losing some info:

| MT System:     | The Tunisian embassy was hit by **bombs** |
| Reference:     | The embassy of Tunisia was hit by **firebombs** |

• Padó et al., 2009
  • Checking if the translation paraphrases the reference using entailment features
• Both are after semantic equivalence or entailment

• MT can be seen as cross-lingual entailment (paraphrasing)
  ⭐ TE definition doesn’t require being monolingual!

• TE (paraphrasing) can be viewed as monolingual translation
TE and MT – it’s already happening

... even if the term “entailment” is not always used...

• MT technology for monolingual tasks
  • Text Simplification via MT (Specia, 2010)
    ★ They ARE actually generating entailed sentences: Original ⇒ simplified

• Monolingual paraphrases used to improve MT (Callison-Burch, and more later)
As just described for MT-evaluation, *directionally-entailed* (more general) translations are sometimes useful.

Loss of information justified in order to:
- Address unknown words
- Simplify complex source structures

Acceptable translations produced, coverage increased
- Measuring information loss is a remaining challenge
A combined TE-MT process

S (Source) The Tunisian embassy in Switzerland was hit by firebombs

T (Target) השגרירות התוניסאית בשוויץ הותקפה

Reference שגרירויות תוניסיה בשווייץ הותקפה בבקבוקי תבערה
Unified view requires entailment *generation*

- Entailment used so far mostly for recognition

- A generative approach for entailment: transformations
  - Generating *entailed consequents*

- Utilizing various types of knowledge (*entailment rules*)
  - Lexical: synonyms, hypernyms (*hit* ⇒ *attack*)
  - Template-based: *X was hit by Y* ⇒ *X was attacked*
  - Syntactic: passive to active

- E.g. BIUTEE (Bar-Ilan University Textual Entailment Engine)
  - Interesting ties to syntax-based SMT techniques
Text: Children like candies
Rules: children ⇔ kids ; like ⇔ enjoy ; candies ⇔ sweets

Consequents:
- Kids like candies
- Kids enjoy candies
- Children like sweets
- …

$2^3$ alternatives!

• We need a packed representation (as in MT, parsing, …)
- **Compact Forest** (Bar-Haim et al., EMNLP-2009)
  - A compact representation of consequents, via hyperedges

- Complexity reduction (typically) from exponential to linear
Intermediate summary

- Entailment and MT are conceptually inter-related
  - seeking equivalence or entailment within and across languages

- MT-technology may be valuable for entailment modeling

- Prospects for integrating entailment in the MT flow
  - First steps presented next...
TE in MT – first steps

Task: Replacing unknown words (OOV) with entailed ones

- ACL-09
  - Showing entailment contribution to MT

- EAMT-10
  - Integration into standard SMT workflow
Addressing OOV via source-language entailment information

(Mirkin et. al, ACL-09)
Motivation – unknown terms

- MT systems frequently encounter terms they are unable to translate - *unknown terms* (OOV)

- Particularly common for:
  - Language-pairs for which parallel corpora are scarce
  - Different training-test domains

⇒ poor translation

**Goal:** improve translation of texts with unknown terms
  ➢ through entailment-based approach
Handling unknown terms – baseline approaches

- Translating to French:
  "Cisco filed a lawsuit against Apple for patent violation"

Baseline approaches:

- Leaving the unknown terms **untranslated**
  "Cisco lawsuit filed une contre Apple pour violation de brevet"

- **Omitting** the unknown terms
  "Un Cisco contre Apple pour violation de brevet"
  ("A Cisco against Apple for...")
Handling unknown terms – paraphrasing

- Translating to French:
  “Cisco filed a lawsuit against Apple for patent violation”

Paraphrasing (Callison-Burch et al., 2006)

- Translating a known paraphrase instead of the original term
- E.g.: file a lawsuit ⇔ sue
  Implicitly translating: Cisco sued Apple for patent violation

- Callison-Burch et al.’s implementation:
  - Requires multilingual corpora
  - Ambiguity is handled by the SMT-standard target LM
Handling unknown terms - textual entailment

• When paraphrases not available, generate source entailments

• E.g.: file a lawsuit $\Rightarrow$ accuse
  
  Cisco *filed a lawsuit* against Apple for patent violation $\Rightarrow$
  
  Cisco *accused* Apple for patent violation
  
  • Improves coverage, still producing useful translations

• Rules are context dependent
  
  • Verify rule application with context models

• Use monolingual source-language Information:
  
  • Monolingual resources & methods are more abundant
  
  • Better suited for directional rules
Textual entailment for MT – input & output

**Input**
- A source text $s$ with one or more unknown terms
- A monolingual resource of entailment rules
- Pruning parameter $k$

**Output**
A translation of either (in order of preference)

1. a paraphrase of $s$
   OR
2. a text entailed by $s$
   OR
3. $s$ with unknown terms left as is
Textual entailment for SMT – method (brief)

Cisco sued Apple

unknown

Cisco processed Apple
Cisco accused Apple

Pruning: k=2
Experimental setting

• **SMT system**: Matrax (Simard et al., 2005)

• **Corpora** (from the shared translation task in WMT-2008):
  - **Training**: Europarl 1M English-French sentences
  - **Test**: ~2,500 News English sentences with unknown terms

• **Entailment rules resource**: WordNet 3.0
  - **Paraphrases**: Synonyms (e.g. provoke ⇔ evoke)
  - **TE**: adding directional entailments: Hypernyms (provoke ⇒ cause)

• **Evaluation**:
  - Manual: annotators marking each translation as acceptable or not
  - Automatic: BLEU, Meteor
### Manual evaluation results

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision (%)</th>
<th>Coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Src</td>
<td>Tgt</td>
</tr>
<tr>
<td>1</td>
<td>–</td>
<td>SMT</td>
</tr>
<tr>
<td>2</td>
<td>NB</td>
<td>SMT</td>
</tr>
<tr>
<td>3</td>
<td>LSA</td>
<td>SMT</td>
</tr>
<tr>
<td>4</td>
<td>NB</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>FREQ</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>RAND</td>
<td>–</td>
</tr>
</tbody>
</table>

- **Target-only model**
- **Source-Target models**
- **Source-only models**

#### Baselines

- **TE vs. Paraphrases:** substantial coverage increase
  - with just a little decrease in precision
- **Src-Tgt models (2-3)** comparable to tgt-only (1), but more efficient
- **Top models outperform the baselines**
Comparison to previous approach

- Comparison to: Callison-Burch et al., 2006 (CB)
  - Phrase table augmentation using Europarl parallel corpora
  - Manual evaluation (150 sentences): acceptance and preference

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision (%)</th>
<th>Coverage (%)</th>
<th>Better (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE</td>
<td>85.3</td>
<td>56.2</td>
<td>72.7</td>
</tr>
<tr>
<td>CB</td>
<td>85.3</td>
<td>24.2</td>
<td>12.7</td>
</tr>
</tbody>
</table>
Conclusions for first model

• A new approach for handling unknown terms in MT

• First application of TE to MT for improving translation quality

• Translation improved through novel components:
  • Monolingual (source-language) resources
  • Directional entailment relationships

• Next step:
  • Better integration into the standard SMT process
Integrating entailment-based replacements into the SMT workflow

(Aziz et. Al, EAMT-10)
• Casting the selection of entailment-based replacements as:
  • A learning problem
  • Active learning (see the paper)
  • Based on human annotations
    • Automatic metrics are unsuitable for semantic modifications
  • With the entailment model tightly integrated into the phrase-based SMT decoder
The mayor was *attacked* by the press
Le maire a été *attacked* par la presse

**Entailment rules/paraphrases**

- attacked ⇔ accused
- attacked ⇔ hit

- (attacked, accused)
- (attacked, touché)

**Phrase pairs (biphrases)**

- (mayor, maire)
- (press, presse)
- (attacked, ?)

- (accused, accusé)
- (hit, touché)

**Static biphrases**

**Dynamic biphrases**
The integrated model

- **Original model**
  \[ \text{argmax}_{(a,t)} \Lambda \cdot G(s,t,a) \]

- **Integrated model**
  \[ \text{argmax}_{(a,t)} \Lambda \cdot G(s,t,a) + M \cdot H(s,t,a) \]
  - standard "static" features
  - dynamic features

- **Dynamic features**
  - Representing properties of the replacements
    - Depend on **the current context**
  - Can use **test domain** source-language monolingual information
  - Only for the dynamic biphrases
    - Avoiding bias on "regular" (non-OOV) sentences
## Entailment features

<table>
<thead>
<tr>
<th></th>
<th>$DSim$</th>
<th>$CSim$</th>
<th>$InfoLoss$</th>
</tr>
</thead>
<tbody>
<tr>
<td>attacked ⇔ accused</td>
<td>-3.1</td>
<td>-0.3</td>
<td>-0.4</td>
</tr>
<tr>
<td>attacked ⇔ hit</td>
<td>-5.2</td>
<td>-7.2</td>
<td>-0.5</td>
</tr>
</tbody>
</table>

## Biphrase features

<table>
<thead>
<tr>
<th>source</th>
<th>target</th>
<th>static features</th>
<th>dynamic features</th>
</tr>
</thead>
<tbody>
<tr>
<td>mayor</td>
<td>maire</td>
<td>-0.1</td>
<td>0</td>
</tr>
<tr>
<td>press</td>
<td>presse</td>
<td>-1.5</td>
<td>0</td>
</tr>
<tr>
<td>accused</td>
<td>accusé</td>
<td>-1.6</td>
<td>0</td>
</tr>
<tr>
<td>hit</td>
<td>touché</td>
<td>-0.9</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>static biphrases</th>
<th>dynamic biphrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>attacked accusé</td>
<td>attacked touché</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results

- **Setting:**
  - Baseline SMT system: MATRAX, 1M Europarl sents
  - OOV model tuning: WMT-09 News Commentary, 1000 sents
  - Evaluation set: 500 OOV sents

<table>
<thead>
<tr>
<th>System</th>
<th>Avg Rank</th>
<th>Best</th>
<th>Acceptance</th>
</tr>
</thead>
<tbody>
<tr>
<td>OOV-Human'</td>
<td>2.274</td>
<td>0.6258</td>
<td>0.7002</td>
</tr>
<tr>
<td>Mirkin09</td>
<td>2.736</td>
<td>0.5172</td>
<td>0.5822</td>
</tr>
<tr>
<td>OOV-MERT</td>
<td>3.153</td>
<td>0.4024</td>
<td>0.4849</td>
</tr>
<tr>
<td>SMT-baseline</td>
<td>3.998</td>
<td>0.1549</td>
<td>0.2918</td>
</tr>
<tr>
<td>Marton09</td>
<td>4.107</td>
<td>0.1690</td>
<td>0.2495</td>
</tr>
</tbody>
</table>
Learning an OOV expert for SMT - summary

• An entailment-based integrated OOV model
  • Dynamic biphrases generated for OOV words
    • Based on entailment rules
  • Assigned with dynamic features
    • Weights learned via human annotation, in active learning scheme
  • Improves SMT performance

• An instance of a more general task:
  Learning an Expert for SMT
TE & MT: Conclusions
Conclusion & future work

• **Unified view of MT and TE**
• **TE information is useful for improving MT**
  - and can be dynamically integrated into standard SMT architecture

**Future work**

• **Improving source-language entailment models**
  - More types of entailment rules
  - Improved context models

• **Additional stages of the unified MT/TE vision**
  - Target language entailments, semantic MT evaluation

• **Quantifying information loss for directional entailment**
Thank you!

Questions ?