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

Outline

- 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

Textual Entailment (TE)

- A generic framework for applied semantic inference
- Core task: *Can the meaning of a target textual assertion (hypothesis, H) be inferred from a given text (T)?*

H *The Tunisian embassy in Switzerland was attacked*

T *Fire bombs were thrown at the Tunisian embassy in Bern*

- In this case: ***T entails H (T ⇒ H)***

- Paraphrasing is bi-directional entailment

T’ *The embassy of Tunisia in Bern was hit by fire bombs*

T *Fire bombs were thrown at the Tunisian embassy in Bern*

- ***T and T’ mutually entail each other (paraphrases) (T ⇔ T’)***

Reducing applications' inferences to entailment

- **Question Answering**

Question	**Expected answer template**
Who founded Wikileaks?	X founded Wikileaks

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

Entailment within Intelligent Tutoring (Nielsen et. Al, 2007)

- **The task**
  - Given student's textual answer to a system's question – asses 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

- 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

TE and MT – a unified view

- 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 $\Rightarrow$ simplified
- Monolingual paraphrases used to improve MT (Callison-Burch, and more later)

Directional Entailment for MT (our work)

- 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**

Reference

שגרירות 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

As usual – generation efficiency required

Text: *Children like candies*

Rules: *children* ⇔ *kids* ; *like* ⇔ *enjoy* ; *candies* ⇔ *sweets*

Consequents:

Kids like candies

Kids enjoy candies

Children like sweets

…

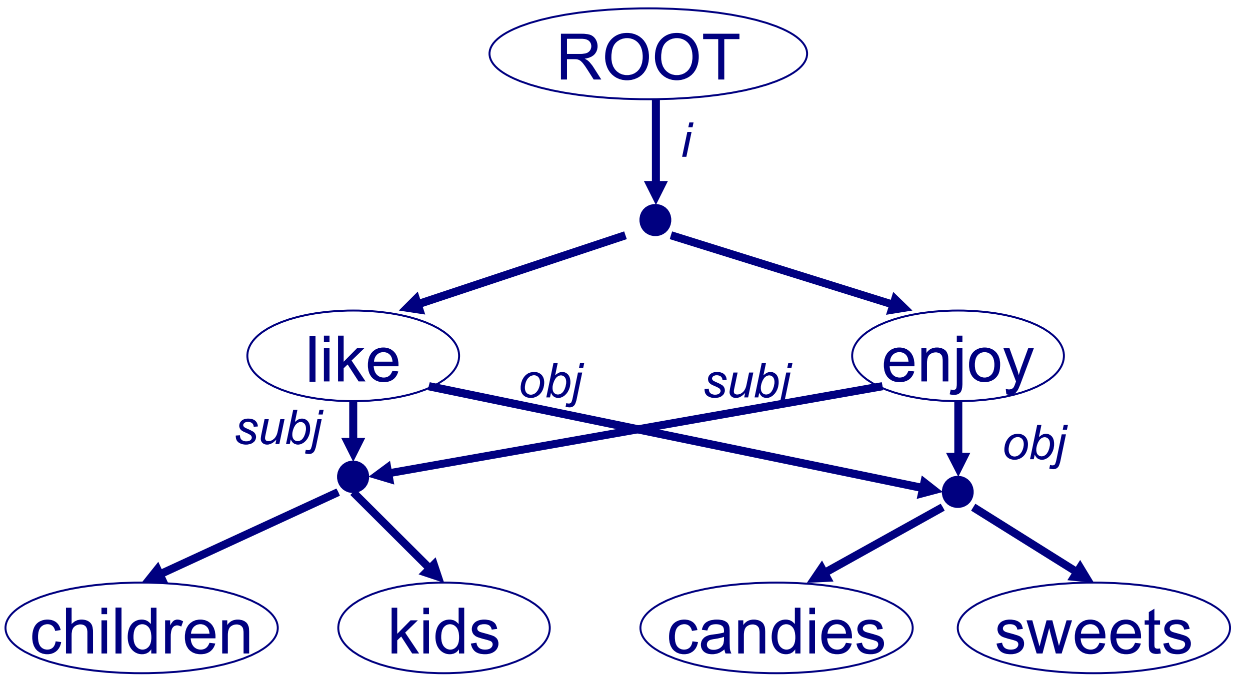
2^3 alternatives!

Exponential explosion!

- We need a packed representation (as in MT, parsing, …)

Children and sweets – the compact version

- *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"*

unknown

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"*

unknown

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*

Textual Entailment for MT

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)

unknown

Cisco sued Apple

Not in phrase table

Source context-model scores

Target-model scores

Source-Target scores

~~sued ⇔ litigated~~
sued ⇔ processed
sued ⇒ accused
sued ⇒ challenged
...

-
0.3
0.6
0.1
...

Cisco processed Apple
Cisco accused Apple

Cisco a mixé une pomme
Cisco a accusé Apple

0.5
0.4

0.15
0.24

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*)

Paraph ⊆ TE

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

Manual evaluation results

Model			Precision (%)		Coverage (%)	
	Src	Tgt	PARAPH.	TE	PARAPH.	TE
1	–	SMT				
2	NB	SMT				
3	LSA	SMT				
4	NB	–				
5	FREQ	–				
6	RAND	–				

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**

Model	Precision (%)	Coverage (%)	Better (%)
TE	85.3	56.2	72.7
CB	85.3	24.2	12.7

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)

OOV replacement as a learning problem

- 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

OOV words and paraphrases/entailment

phrase pairs
(*biphrases*)

The mayor was *attacked* by the press
Le maire a été *attacked* par la presse

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

Entailment rules/paraphrases

attacked ⇔ accused
attacked ⇔ hit

(accused, accusé)
(hit, touché)

Static biphrases

Dynamic biphrases

(attacked, accusé)
(attacked, touché)

Le maire a été accusé par la presse

The integrated model

- Original model

$$\text{argmax}_{(a,t)} \Lambda \cdot G\,(s,t,a)$$

- Integrated model

$$\text{argmax}_{(a,t)} \Lambda \cdot \underbrace{G\,(s,t,a)}_{\text{standard ``static'' features}} + M \cdot \underbrace{H\,(s,t,a)}_{\text{dynamic features}}$$

- 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

Features

attacked ⇔ accused
attacked ⇔ hit

(accused, accusé)
(hit, touché)

(attacked, accusé)
(attacked, touché)

Entailment features

	$DSim$	$CSim$	$InfoLoss$
attacked ⇔ accused	-3.1	-0.3	-0.4
attacked ⇔ hit	-5.2	-7.2	-0.5

Biphrase features

	source	target	static features		dynamic features		
static biphrases	mayor	maire	-0.1	-0.1	0	0	0
	press	presse	-1.5	-0.7	0	0	0
	accused	accusé	-1.6	-1.2	0	0	0
	hit	touché	-0.9	-0.5	0	0	0
dynamic biphrases	attacked	accusé	0	0	-3.1	-0.3	-0.4
	attacked	touché	0	0	-5.2	-7.2	-0.5

Results

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

System	Avg Rank	Best	Acceptance
OOV-Human'	2.274	0.6258	0.7002
Mirkin09	2.736	0.5172	0.5822
OOV-MERT	3.153	0.4024	0.4849
SMT-baseline	3.998	0.1549	0.2918
Marton09	4.107	0.1690	0.2495

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 ?