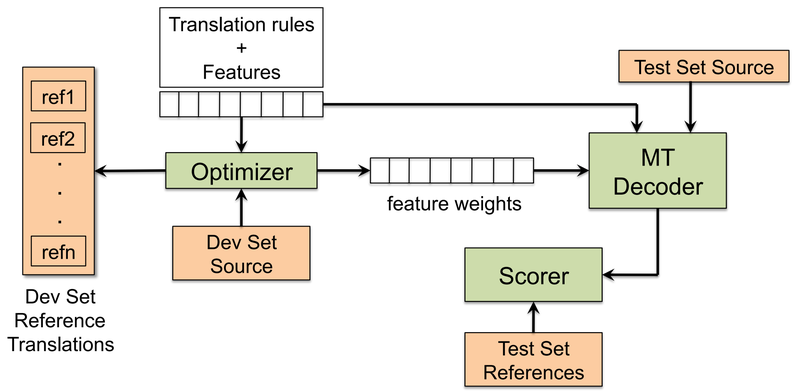
Applying Automatically Generated Semantic Knowledge
A Case Study in Machine Translation

Nitin Madnani Philip Resnik Bonnie Dorr Richard Schwartz
University of Maryland, College Park **BBN Technologies**

Abstract

We discuss how we apply automatically generated semantic knowledge to benefit statistical machine translation (SMT). Currently, all statistical machine translation systems rely heavily on memorizing translations of phrases. Some systems attempt to go further and generalize these learned phrase translations into templates using empirically derived information about word alignments and a small amount of syntactic information, if at all. There are several issues in an SMT pipeline that could be addressed by the application of semantic knowledge, if such knowledge were easily available. One such issue, an important one, is that of reference sparsity that is our focus in this work.

1. A Typical SMT System



- Translation rules and features are learned automatically from bi-text.
- The MT Decoder is essentially a CKY-parser with beam search.

2. Reference Sparsity

- **Fundamental Problem**: No such thing as the correct translation. A source sentence can be translated into the target language in many valid ways.

- This requires *multiple human-authored* reference translations during development to better tune the feature weights relative to a given translation quality metric.

- This requirement is difficult to satisfy:

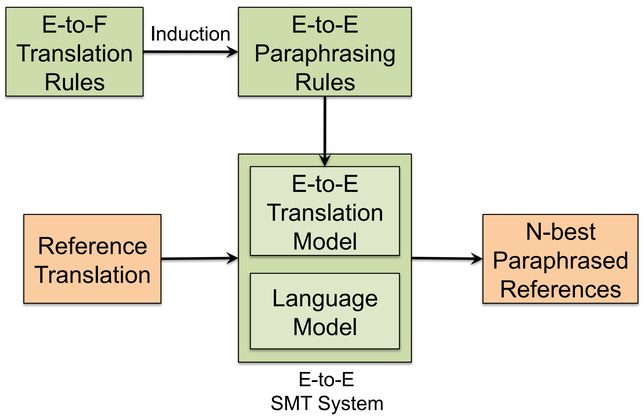
  - Almost all current and new MT research datasets are provided with only a single reference translation to save time and expense.

  - Obtaining such translations in rapid development, low-density language scenarios is impractical due to limited availability of translators.

3. Our Solution: Automatic Paraphrasing as Translation

- Monolingual knowledge can be considered inherent in parallel bilingual corpora and looking at two languages in parallel provides a way to "triangulate" on semantics.

- Specifically, we extend a well-known *pivot-based* methodology of extracting (approximately) semantically equivalent phrasal correspondences to build a full-sentence paraphrasing model.



Using the E-to-E translator (paraphraser) to create artificial reference translations from a single given human reference

Pivoted Paraphrase Rule Induction (F: *any* foreign language)

- Use E-to-F translation rules extracted during regular SMT training

- For each pair of rules with a common F, pivot on F to induce an E-to-E paraphrastic pair (Bannard et al., 2005)

- $Count_{E1\text{-}to\text{-}E2} = Count_{E1\text{-}to\text{-}F} * Count_{E2\text{-}to\text{-}F}$

- For each induced rule, compute the following as features: $p(E_1, E_2)$, $p(E_2 | E_1)$, $p(E_2 | E_1)$ and $||E_2||$

X_1 建 X_2 ||| X_1 to build X_2 ||| 2.0
X_1 建 X_2 ||| X_1 to construct X_2 ||| 1.0
X_1 建 X_2 ||| X_1 to establish X_2 ||| 3.0

↓

X_1 to build X_2 ||| X_1 to construct X_2 ||| 2.0
X_1 to build X_2 ||| X_1 to establish X_2 ||| 6.0
X_1 to construct X_2 ||| X_1 to establish X_2 ||| 3.0
. . . . . .

Example of Rule Induction (F: Chinese)

4. Examples

Pivot Language = French

We must bear in mind the community as a whole. (original)
We must remember the wider community. (paraphrase)

Thirdly, the implications of enlargement for the union's regional policy cannot be overlooked.
Finally, the impact of enlargement for EU regional policy cannot be ignored.

They should be better coordinated and more effective.
They should improve the coordination and efficacy.

Pivot Language = Chinese

He was now preparing a speech concerning the US policy for the upcoming World Economic Forum.
He was now ready to talk with regard to the US policies for the forthcoming International Economic Forum.

France sent its proposal in the form of a "non-official paper".
French transmits its recommendations to serve as a "non-official document".

Alcatel added that the company's whole year earnings would be announced on February 4.
Alcatel said that the company's total revenues would be released on February 4.

5. Results

Tuning References	Newswire		Web	
	BLEU	TER	BLEU	TER
1H	37.65	56.39	15.17	70.32
1H+1P	39.32*	54.69	15.92*	69.94

*statistically significant at $p < 0.05$

We obtain statistically significant improvements in BLEU and TER scores for Chinese-English translation of standard test sets in Newswire and Web genres when using even a single additional artificial reference for optimization.

(1H=Tuning with 1 human reference, 1H+1P=Tuning with the human reference and its paraphrase. Higher BLEU scores are better and lower TER scores are better.)