The 1st Workshop on Semantics-Driven Statistical Machine Translation (S^{2}MT 2015)

Deyi Xiong, Kevin Duh,
Christian Hardmeier and Roberto Navigli

ACL-IJCNLP 2015 Workshop
July 30, 2015
Beijing, China

Opening Remarks

Why

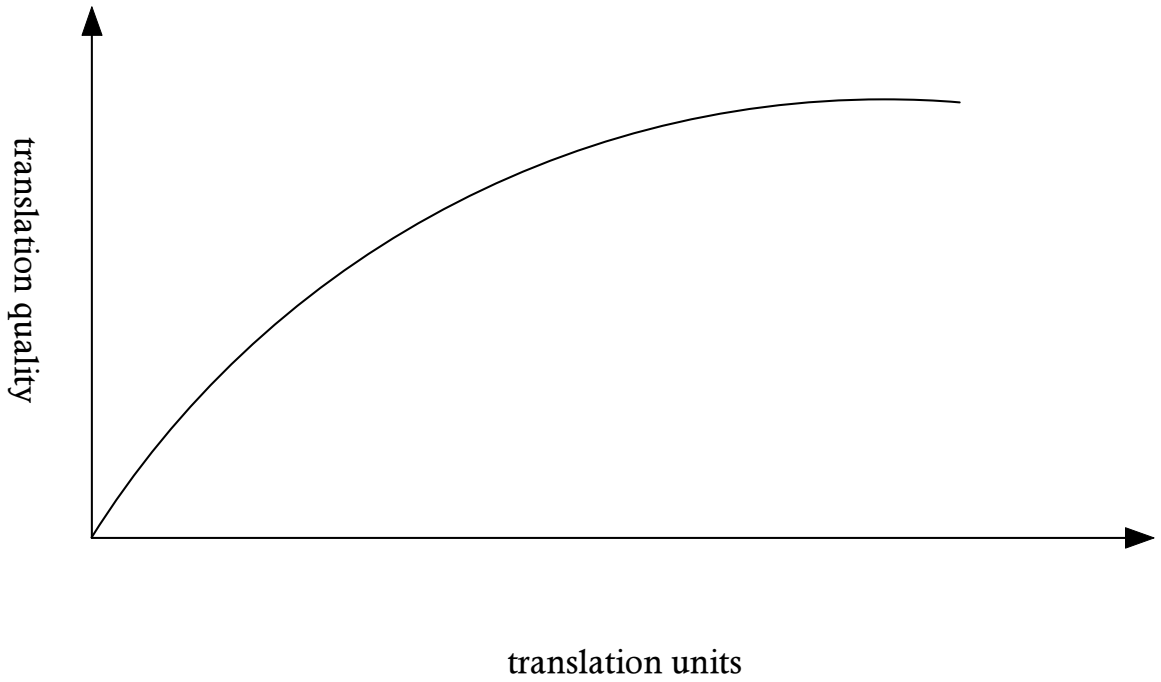
What

Goals

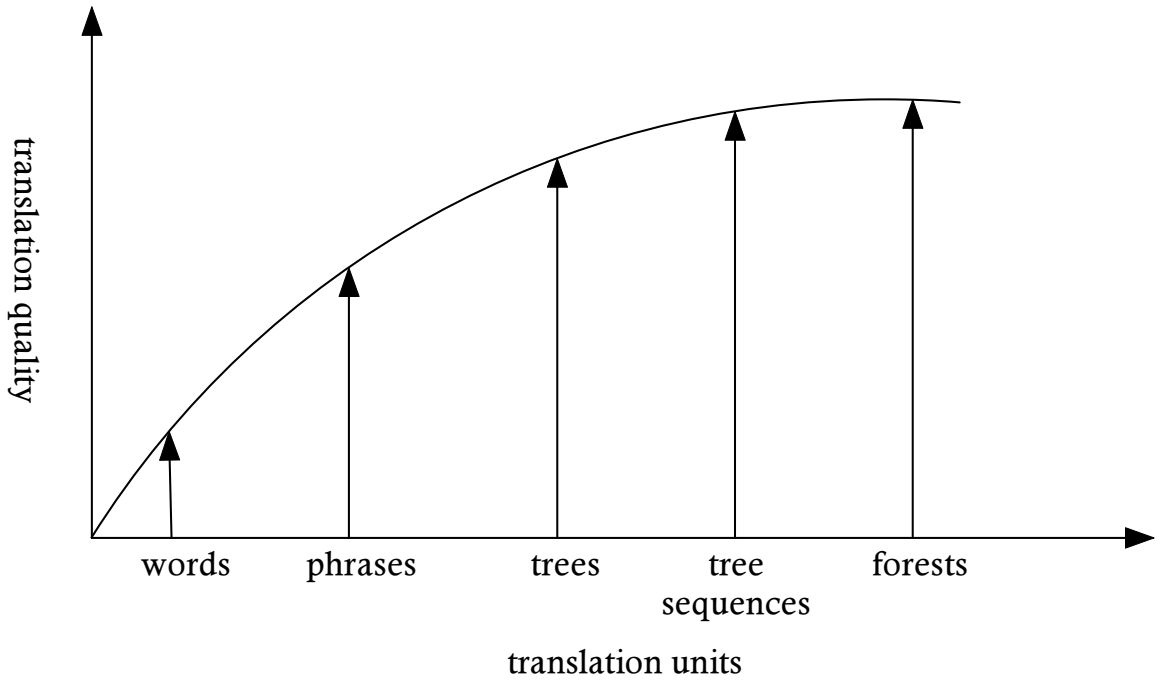
Program

S^{2}MT: Why

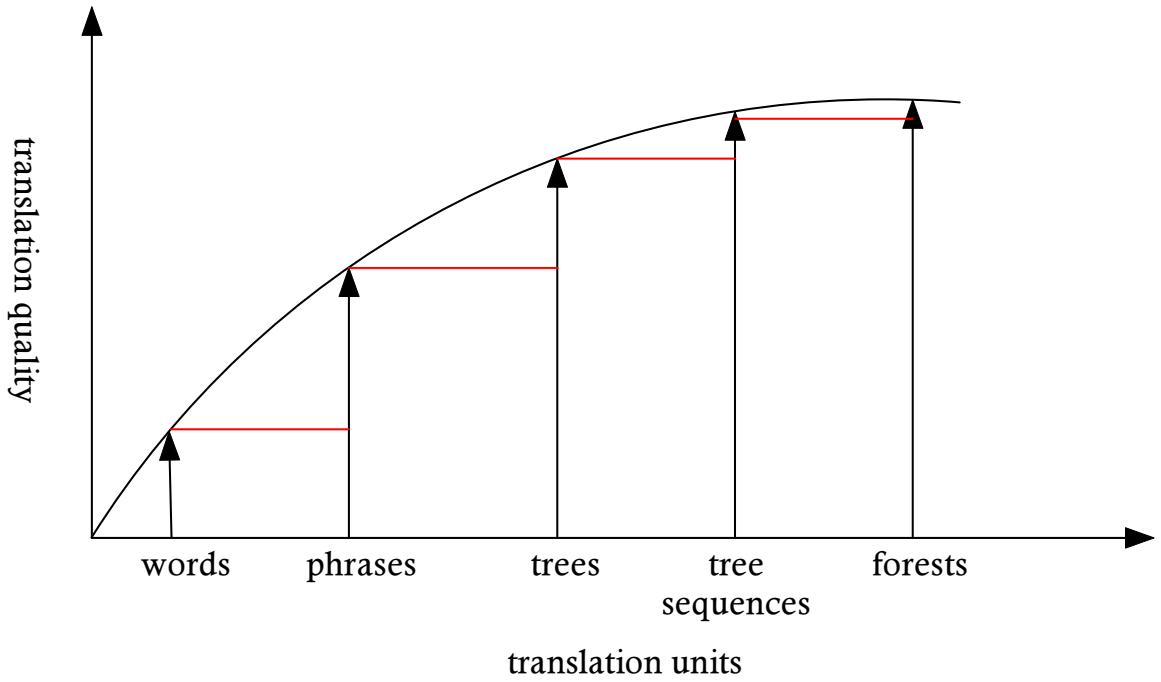
SMT: Current Dilemma



SMT: Current Dilemma



SMT: Current Dilemma



Reasons behind the Dilemma

Lack of "deep" understanding

[Nirenberg:89]:

> *In practice, those MT researchers who believe in translating without 'deep' understanding of the SLG text tend to prefer the transfer paradigm. The price they have to pay for avoiding meaning analysis is the need for an extra step in the translation process, namely, postediting.*

"Know-nothing"

[Knight:95]:

> *"know-nothing" statistical MT... "know-it-all" knowledge-based MT...*

Semantics-Driven Statistical Machine Translation

- Substantial progresses in computational semantics for meaning analysis at different levels
- Many open large-scale (multilingual) knowledge bases, e.g., YAGO, DBpedia, BabelNet

S^{2}MT: make SMT "know-something" at different linguistic levels

It is time for SMT researchers to collaborate with colleagues from the community of semantics and to harvest further significant improvements in translation quality!

S^2MT: What

Syntax-Based SMT vs. Semantics-Driven SMT

	word	phrase	sentence
Syntax	PoS tags	syntax categories	parse trees
SBSMT	PoS-enhanced SMT	SC-enhanced SMT	various tree-based formalisms
Semantics	word senses	semantic compositions	deep semantic representations
S^{2}MT	WSD/WSI for SMT	phrasal semantics for SMT	ongoing

- PoS/SC-enhanced SMT: using PoS tags or syntactic categories to improve reordering/translation or formally syntax-based SMT
- tree-based formalisms: {tree-to-string/tree, string-to-tree, tree sequence, forest} based SMT

Syntax-Based SMT vs. Semantics-Driven SMT

- Semantics-driven SMT is still in its baby stage
  - S^{2}MT does not dominate SMT
- Semantics can provide more for SMT
  - discourse semantics: cross-sentence semantic relations
  - external semantic knowledge: entity relations, ontology and so on

S^{2}MT: Goals

Three Fundamental Goals

- Form a community
- Build resources and tools
- Organize international evaluations

S^{2}MT Community

- A combination of researchers from semantics and SMT
- A dedicated workshop or special interest group
- Vision: shift the interest of the SMT community from syntax to semantics

Resources and Tools

- Develop an open source package that provides a unified framework for semantics-driven SMT, just like Moses for phrase- and syntax-based SMT
- Build large-scale semantically annotated bilingual data for SMT
- Provide various open semantic tools especially designed for SMT

Evaluations

- SMT evaluations:
  - NIST: constrained and unconstrained evaluations, very helpful for phrase- and syntax-based SMT
  - IWSLT: spoken language translation
  - WMT: evaluation tasks for translation, metrics and quality estimation
- Semantic evaluations
  - Senseval: initiated for WSD evaluation
  - SemEval: a number of shared tasks designed for multiple-level multilingual semantic analysis
- S^{2}MTeval: combination of SMT evaluation and semantic evaluation.

S^{2}MT: Program

2 Oral Presentations: Morning Session

10:00–10:30 *Round trips with meaning stopovers*
Alastair Butler

12:00–12:30 *Conceptual Annotations Preserve Structure Across Translations: A French-English Case Study*
Elior Sulem, Omri Abend and Ari Rappoport

2 Oral Presentations: Afternoon Session

14:30–15:00 *Integrating Case Frame into Japanese to Chinese Hierarchical Phrase-based Translation Model*
Jinan Xu, Jiangming Liu, Yufeng Chen, Yujie Zhang, Fang Ming and Shaotong Li

15:00–15:30 *A Discriminative Model for Semantics-to-String Translation*
Aleš Tamchyna, Chris Quirk and Michel Galley

2 Keynote Speeches: Morning Session

9:00–10:00 *Semantic Parsing as, via, and for Machine Translation*
Percy Liang (Stanford University)

11:00–12:00 *Learning Multilingual Semantics from Big Data on the Web*
Gerard de Melo (Tsinghua University)

2 Keynote Speeches: Afternoon Session

13:30–14:30 *Sequence to Sequence Learning for Language Understanding*
Quoc V. Le (Google)

16:00–17:00 *Machine Translation and Deep Language Engineering Approaches*
António Branco (University of Lisbon)

Panel

- Topic: *Semantics and Statistical Machine Translation: Gaps and Challenges*
- Panel Chair: Chris Quirk
- Invited Panel Speakers: Eduard Hovy, Percy Liang, António Branco, Quoc V. Le

Thanks

- Program Committee
- Keynote Speakers: António Branco, Quoc V. Le, Percy Liang, Gerard de Melo
- Invited Panel Speakers: Chris Quirk, Eduard Hovy, Percy Liang, António Branco, Quoc V. Le
- Authors
- Sponsors: NiuTrans, NSFC