# Applying Automatically Generated Semantic Knowledge
# A Case Study in Machine Translation

Nitin Madnani[§], Philip Resnik[§], Bonnie Dorr[§] & Richard Schwartz[†]
[§]Institute for Advanced Computer Studies, University of Maryland, College Park
[†]BBN Technologies
{nmadnani,resnik,bonnie}@umiacs.umd.edu schwartz@bbn.com

In this paper, we discuss how we apply automatically generated semantic knowledge to benefit statistical machine translation (SMT). Currently, almost 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 a 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. The fundamental problem that translation systems have to face is that there is no such thing as *the* correct translation for any sentence. In fact, any given source sentence can often be translated into the target language in many valid ways. Since there can be many "correct answers," almost all models employed by SMT systems require, in addition to a large bitext, a held-out development set comprised of *multiple* high-quality, human-authored reference translations in the target language in order to tune their parameters relative to a translation quality metric.[1] There are several reasons that this requirement is not an easy one to satisfy. First, with a few exceptions—notably NIST's annual MT evaluations—most new MT research data sets are provided with only a single reference translation. Second, obtaining multiple reference translations in rapid development, low-density source language scenarios (e.g. (Oard, 2003)) is likely to be severely limited (or made entirely impractical) by limitations of time, cost, and ready availability of qualified translators.

| |
|---|
| alcatel added that the company's whole year earnings would be announced on february 4.<br>alcatel said that the company's total annual revenues would be released on february 4. |
| he was now preparing a speech concerning the us policy for the upcoming world economic forum.<br>he was now ready to talk with regard to the us policies for the forthcoming international economic forum. |
| tibet has entered an excellent phase of political stability, ethnic unity and people living in peace.<br>tibetans have come to cordial political stability, national unity and lived in harmony. |
| its ocean and blue-sky scenery and the mediterranean climate make it world's famous scenic spot.<br>its harbour and blue-sky appearance and the border situation decided it world's renowned tourist attraction. |

Table 1: Example paraphrases with Chinese as the pivot language. The first sentence in each row is the original reference translation and the second is its sentential paraphrase generated by our paraphrasing model. Examples were deliberately chosen to illustrate the range of quality in the paraphrases we obtain.

In our work, we leverage the thesis that monolingual semantic knowledge can be considered inherent in parallel bilingual corpora, i.e., looking at two languages in parallel translation provides a way to "triangulate" on semantics without having to commit to overt semantic representations (Resnik, 2004). Specifically, we extend a well-known pivot-based methodology (Bannard and Callison-Burch, 2005) of extracting phrasal correspondences that are approximately semantically equivalent and build a full-sentence paraphrasing model (Madnani et al., 2007; Madnani et al., 2008) that we then apply to a *single* good reference translation for each sentence in an MT development set. By doing this, we create additional "reference" translations that can be used during the tuning process. Some examples of these paraphrased references produced by our model are shown in Table 1. Table 2 shows that using even one additional paraphrased reference leads

[1]The most commonly used metric for this purpose is BLEU (Papineni et al., 2002).

to significant improvements over using just the original single reference translation in multiple genres, as measured by both BLEU and TER (Snover et al., 2006). Our analysis also suggests that it is necessary to invest in four or more *human* translations in order to significantly improve on a single translation augmented by paraphrases.

| Tuning refs | Newswire | | Web | |
|---|---|---|---|---|
| | BLEU | TER | BLEU | TER |
| 1H | 37.65 | 56.39 | 15.17 | 70.32 |
| 1H+1P | 39.32 | 54.39 | 15.92 | 69.94 |

Table 2: BLEU and TER scores for Chinese-English translation of standard test sets in Newswire and Web genres. 1H=Tuning with 1 human reference, 1H+1P=Tuning with the human reference *and* its paraphrase. Lower TER scores are better.

However, our results point in an even more ambitious direction: doing away entirely with any human translations beyond those already a part of the bitext already expected by statistical MT systems. If the quality of the translations in the training set are good enough—or if a high quality subset can be identified—then our paraphrasing techniques may suffice to obtain the target-language diversity needed to tune statistical MT systems effectively. Experimentation of this kind is a priority for our future work.

## References


Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. In *Proceedings of ACL*.

Nitin Madnani, Necip Fazil Ayan, Philip Resnik, and Bonnie J. Dorr. 2007. Using paraphrases for parameter tuning in statistical machine translation. In *Proceedings of the Second ACL Workshop on Statistical Machine Translation*.

Nitin Madnani, Philip Resnik, Bonnie J. Dorr, and Richard Schwartz. 2008. Are multiple reference translations necessary? investigating the value of paraphrased reference translations in parameter optimization. In *Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas*.

D. W. Oard. 2003. The surprise langauge exercises. *ACM Transactions on Asian Language Information Processing*.

K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of ACL*.

Philip Resnik. 2004. Exploiting hidden meanings: Using bilingual text for monolingual annotation. *Lecture Notes in Computer Science 2945: Computational Linguistics and Intelligent Text Processing*, pages 283–299.

M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of AMTA*.


## Acknowledgments


This work was supported, in part, by BBN under DARPA/IPTO contract HR0011-06-C-0022 and IBM under HR0011-06-2-001. Any opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the view of DARPA.