Abstract

In this paper, it is described how an Assamese sentence is translated to English using statistical machine translation. Statistical Machine Translation is the paradigm where translations from source to target language are based on statistical models. Moses is used as a platform for Statistical Machine Translation. GIZA++ is also used for word-alignment and IRSTLM for language model training. A Transliteration model is also integrated into the system to deal with out of vocabulary (OOV) words.

References

- Marian Olteanu et al, &quot;Phramer: Open Source Statistical Phrase-Based Translator&quot;, In the Proceedings of the Workshop of Statistical Machine Translation, June
2006, pp. 146-149.
- Peter F. Brown et al., "A Statistical Approach to Machine Translation";
- Philipp Koehn et al., "Statistical Phrase-Based Translation";
  In the Proceedings of HLT-NAALC, May-June 2003, pp. 48-54.
- Philipp Koehn, "Pharaoh: A beam search decoder for phrase based statistical
  machine translation models"; In the proceedings of AMTA, 2004.
- Philipp Koehn et al., "Moses: Open Source Toolkit for Statistical Machine
  Translation"; In the Proceedings of the ACL, June 2007, pp. 177-180.
- Sanjay Kumar Dwivedi and Pramod Premdas Sukhadeve, "Machine Translation
  System in Indian Perspectives"; Journal of Computer Science, Volume 6, Issue 10, pp.
  1111-1116.
- Md. Zahurul Islam, "English to Bangla Statistical Machine Translation";
- Philipp Koehn, "Noun Phrase Translation"; PhD Thesis, University of
  Southern California, 1993.
- "PSMT"; Available: http://psmt.sourceforge.net/
- "The EGYPT Statistical Machine Translation Toolkit"; Available:

**Index Terms**

Computer Science  
Artificial Intelligence

**Keywords**

Assamese  
English  
Statistical Machine Translation  
Transliteration  
Corpus.
Assamese to English Statistical Machine Translation Integrated with a Transliteration Module