INVITED TALK:
The operation sequence model: Integrating translation and reordering operations in a single left-to-right model

Hinrich Schütze

Chair of Computational Linguistics

Center for Information and Language Processing

University of Munich (LMU)

Oettingenstr 67

80538 München, Germany

We present a novel machine translation model that combines the benefits of phrase-based and n-gram-based statistical machine translation (SMT) and remedies their drawbacks. The model is based on a joint source channel probability model that represents the translation process as a linear sequence of operations. The sequence includes not only translation operations but also reordering operations. As in n-gram-based SMT, the model is: (i) based on minimal translation units, (ii) takes both source and target information into account, (iii) does not make a phrasal independence assumption and (iv) avoids the spurious phrasal segmentation problem. As in phrase-based SMT, the model (i) has the ability to memorize lexical reordering triggers, (ii) builds the search graph dynamically and (iii) decodes with large translation units during search. The model has two important properties. First, it strongly couples reordering and translation; this provides a better reordering mechanism for restricting the position to which a word or phrase can be moved and provides a single framework that handles both short and long distance reorderings effectively. Second, no hard reordering constraint needs to be imposed on the model; phrase-based models must impose such constraints because they model reordering poorly. Using BLEU as a metric of translation accuracy, we found that our system performs significantly better than state-of-the-art phrase-based systems (Moses and Phrasal) and n-gram-based systems (Ncode) on standard translation tasks.