

## Abstract

Probabilistic Synchronous Context-free Grammars (PSCFGs) [Aho and Ullmann, 1969, Wu, 1996] define weighted transduction rules to represent translation and reordering operations. When translation models use features that are defined *locally*, on each rule, there are efficient dynamic programming algorithms to perform translation with these grammars [Kasami, 1965]. In general, the integration of non-local features into the translation model can make translation NP-hard, requiring decoding approximations that limit the impact of these features.

In this thesis, we consider the impact and interaction between two non-local features, the  $n$ -gram language model (LM) and labels on rule nonterminal symbols in the Syntax-Augmented MT (SAMT) grammar [Zollmann and Venugopal, 2006]. While these features do not result in NP-hard search, they would lead to serious increases in wall-clock runtime if naive dynamic programming methods are applied.

We develop novel two-pass algorithms that make strong decoding approximations during a first pass search, generating a hypergraph of sentence spanning translation derivations. In a second pass, we use knowledge about non-local features to explore the hypergraph for alternative, potentially better translations. We use this approach to integrate the  $n$ -gram LM decoding feature as well as a non-local syntactic feature described below.

We then perform a systematic comparison of approaches to evaluate the relative impact of PSCFG methods over a strong phrase-based MT baseline with a focus on the impact of  $n$ -gram LM and syntactic labels. This comparison addresses important questions about the effectiveness of PSCFG methods for a variety of language and resource conditions. We learn that for language pairs that exhibit long distance reordering, PSCFG methods deliver improvements over comparable phrase-based systems and that SAMT labels result in additional small, but consistent improvements even in conjunction with strong  $n$ -gram LMs.

Finally, we propose a novel approach to use nonterminal labels in PSCFG decoding by extending the PSCFG formalism to represent hard label constraints as soft preferences. These preferences are used to compute a new decoding feature that reflects the probability that a derivation is syntactically well formed. This feature mitigates the effect of the commonly applied *maximum a posteriori* (MAP) approximation and can be discriminatively trained in concert with other model features. We report modest improvements in translation quality on a Chinese-to-English translation task.