Non-local features in Syntactic parsing
(Abstract of thesis submitted for completion of MS (by Research) in CSE)
sudheer.kpg08@research.iiit.ac.in

Natural language parsers lie at the core of various natural language processing (NLP) systems such as machine translation (MT), question answering (QA), information extraction/retrieval (IE/IR), etc. Building accurate, wide-coverage parsers has been one of the main goals in NLP research for the last two decades. As a result, there exist today highly accurate parsers based on a variety of approaches not only for English but also for a few other languages. Statistical parsers have proven to be most effective both in terms of coverage and precision. However, statistical parsers whether constituency-based or dependency-based make certain independence assumptions about sentence structure. Simply put, statistical parsers of all hues assume that sentence structure can be factored into smaller sub-structures which can be predicted independently varying only in the type of factorization. Although such assumptions are necessary in order to ensure tractability of parsing algorithms, they are not linguistically tenable since we know that there is a significant amount of interaction among the factored sub-structures. There are a number of linguistic phenomena such as subject-verb agreement, verb argument structure and coreference of noun phrases where information about the linguistic relationship is spread over more than one sub-structure. In other words, in all these cases, the cues for parsing one sub-structure correctly can come from another sub-structure. Most state-of-art statistical parsing models which make independence assumptions about sentence structure fail to capture such non-local phenomena.

In this thesis, I study two approaches to overcome this limitation and make use of non-local features that encode greater contextual information during parsing. The first approach is based on the technique of discriminative reranking which consists of two steps: increasing the width of the search beam to allow more candidate parses and then employing a classifier that can use non-local features to rank the candidate parses and pick the best among them. The second approach is based on the technique of ensemble parsing whereby parsing models with complementary strengths/weaknesses can be combined to obtain best possible parsing performance. In particular, I study the stacking approach to combining parsers at learning time. The explored experimental setup allows for non-local features defined over the output of one (or more) parser(s) to be used while training a graph-based parser. In my experiments on discriminative reranking and ensemble parsing, I build several highly accurate parsers for English which can be directly used in in-house large-scale English-to-Indian language machine translation systems. I combine freely available parsers using the ensemble technique of re-parsing to build the best performing model for dependency parsing of English. This high accuracy dependency parser for English is available under GPL and can be deployed in a wide array of NLP systems. In my experiments on stacking dependency parsers, I build stacked parsing models with different combinations of non-local features for three Indian languages—Hindi, Telugu and Bangla—to study the influence of each feature in improving parsing performance. The accuracies of the best performing of these models are the state-of-art accuracies for parsing these Indian languages. Although I do not come up with a entirely new way of overcoming the limitation of feature locality in statistical parsing frameworks, the insights gained from the studies presented in this thesis can inform efforts aimed at development of contextually rich models of syntactic parsing.