

Abstract

In this thesis, we have developed cognitively motivated computational models for the representation and processing of Bangla morphologically complex words (like, derivationally suffixed, compounds and phrasal verbs) in a computational lexicon. To achieve this, we have performed different psycholinguistic experiments to collect data and explore the interaction between different linguistic factors in the possible representation and processing of words in the human mind. Based on the empirical results we have developed and applied different computational models like, frequency and productivity based models, information theoretic measures, vector space models, semantic compositionality and other distributional semantics techniques to predict our empirical observations.

Through different psycholinguistic experiments, we have observed that the representation of words depends on the activation and deactivation of its constituent bound and/or free morphemes. Such activation and deactivations are primarily modulated by the morphological relatedness, orthographic transparencies and semantic compositionality between the words and their constituents. These observations are made by examining the priming effect between the words and their constituents. Further analysis of the reaction time and error rates reveals several interesting facts such as (a) apart from usage frequency, word length and presence of certain orthographical features also affect the representation and processing of a word, (b) in case of derived words, certain derivational suffixes inherited from Sanskrit, which usually make the derived word orthographically or semantically opaque, do not trigger priming; this indicates that these morphological relations are no longer recognized or internalized by the modern Bangla speakers and (c) apart from morphological relatedness, semantic compositionality between the whole word and its constituents also plays an important role in the representation and processing of Bangla morphologically complex words. These and similar other observations make us believe that understanding the precise nature of the mental representation of morphological processes in Bangla (as well as other Indian languages) is a challenging and potent research area that is very little explored.

Based on the above empirical experiments we have collected a large sample of reaction time data and incrementally developed a number of frequency based, morphological complexity and semantic compositionality based computational models. Finally, we have combined all these individual models together to develop an enhanced parallel lexical activation model that can predict the possible representation and processing of Bangla words in the mental lexicon. This model is then augmented to a lexical representation scheme to decide which words should be listed in the lexicon and which one are to be processed on the fly using a morphological analyzer. A representation scheme for the computational lexicon based on the principles of mental lexicon organization is expected to perform better because its success and failure in processing words are expected to meet the expectations of the end user.

Keywords: Lexical representation scheme, Mental lexicon, Morphologically complex words, Frequency effects, Morphological complexity, Semantic compositionality