

A dependency annotation scheme to extract syntactic features in Indonesian sentences

Budi Irmawati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920534364&lokasi=lokal>

Abstrak

In languages with fixed word orders, syntactic information is useful when solving natural language processing (NLP) problems. In languages like Indonesian, however, which has a relatively free word order, the usefulness of syntactic information has yet to be determined. In this study, a dependency annotation scheme for extracting syntactic features from a sentence is proposed. This annotation scheme adapts the Stanford typed dependency (SD) annotation scheme to cope with such phenomena in the Indonesian language as ellipses, clitics, and non-verb clauses. Later, this adapted annotation scheme is extended in response to the inability to avoid certain ambiguities in assigning heads and relations. The accuracy of these two annotation schemes are then compared, and the usefulness of the extended annotation scheme is assessed using the syntactic features extracted from dependency-annotated sentences in a preposition error correction task. The experimental results indicate that the extended annotation scheme improved the accuracy of a dependency parser, and the error correction task demonstrates that training data using syntactic features obtain better correction than training data that do not use such features, thus lending a positive answer to the research question.