Abstract

Grammar correction tools intended for native speakers of English are common, and those directed towards ESL students are becoming increasing common as well. However, there are apparently no software tools available that provide grammatical feedback to advanced learners of language, individuals who are generating grammatically correct samples of the target language, yet whose output is still distinguishable from that produced by a native speaker. This study explores how such a system might be developed. The focus is on how automatically-generated parses of text can be mined for grammatical features which can then be used in machine-learning algorithms to classify the text as having been generated by a native or nonnative speaker. Using a number of native and nonnative corpora of written English as training and testing data, this study shows that such classification can be done with a high level of accuracy. This study also proposes and describes an interactive system that would take advantage of this classification process to provide the learner with detailed information on which aspects of his or her language mark him or her as a nonnative speaker.