

Annotation of Learner Corpora

Developing a markup scheme

Anna Kosterina
Mohammad Haji-Abdolhosseini
Iowa State University

Date: Tue. Dec. 12, 2006

Time: 3:00 p.m.

Location: 212 Ross Hall

One potentially useful research area in second language acquisition is the study of sequences of learners' development of grammatical knowledge such as question formation, negation, tense, complex clauses, and collocations. Learner corpora provide large amounts data allowing for potentially making quantifiable and fairly reliable observations about learner language use. However before any corpus-based research can find applications in second language education, fundamental methodological issues need to be addressed.

This presentation will report on our work in progress in developing a markup scheme for objectively analyzing language use patterns in English texts written by English learners. We will discuss the design principles guiding our work and some of the challenges that the task presents.