A New Look at Learner Language - the Trinity Lancaster Corpus
Tony McEnery

Does cultural and linguistic background affect learner speech - and if so how? What impact may age have on learner production? Is gender a linguistically important feature when exploring the speech of learners of English? How does learner language production vary by task type? Is learner language different when a learner is leading an interaction as opposed to being led through an interaction by a person who is proficient in the language?

Questions such as these have been addressed regularly in the literature on learner language. However, until recently it was difficult to explore these questions in learner speech. Using a new, multi-million word, corpus being developed at Lancaster University with Trinity College London we can start to address these issues. By exploiting this large, orthographically transcribed, corpus of learner speech, amply provided with plentiful relevant metadata, we can gain fresh insights into learner speech.

In this talk I will overview the construction of the Trinity Lancaster Corpus, discussing the tasks the speakers engaged in and the range of metadata we have available for those speakers. Following from that I will review some initial findings from the corpus. The findings will use a range of metadata to show how, when considered singly and in groups, that metadata can give us answers to questions such as those outlined.