Spoken linguistic data transcription and data mining

The task of orthographic transcription of recordings of oral conversation is a complex one, and reveals a great deal about the nature of spoken language (vis a vis the written language we are used to reading). It foregrounds the non-standardness of language structure, the importance for speech of conversational grammatical phenomena such as discourse markers, as well as the intricacy of how interlocutors cooperatively manage and alternate talk. We have a set of recordings useful for language variation analysis that can be transcribed for Independent Study. You can get 1 ECTS per completely orthographically transcribed recording. See Prof Britain for more details.

The next stage of an analysis of linguistic variation would be to extract relevant tokens of the variable in question from the transcribed data; for example extracting all examples of relative clauses from the transcriptions, and preparing an excel spread sheet containing all the relevant examples, along with relevant background speaker information from the recording. You can get an additional 1 ECTS for extracting *all* examples of 6 linguistic features from a transcribed recording (you would need to listen to the recording with the transcription and not just read the transcription for this task – for several of these variables computerised word searching won't work…): these 5 linguistic variables would be: relative clauses; examples of verbs of obligation (e.g. I have to go; I must go; I have got to go; I gotta go); examples of verbs of possession (e.g. I have a book; I've got a book; I got a book); examples of complementiser presence or absence (e.g. he said that he was Irish v he said he was Irish); examples of the future tense (e.g. I will break; I'll break; I'm going to break; I'm gonna break); contraction in the context of negated auxiliary verbs (e.g. he's not big; he isn't big; he ain't big). There would be a short reading to do first, to help you familiarise yourself with the variables being extracted. See Prof Britain for more details. If you are specifically further interested in the subsequent analysis of the extracted data, for example examining the linguistic factors which shape the choice of variants for these features, please come and see Prof Britain so we can discuss a more tailor-made Independent Studies project that involves further data analysis.