The EVA Corpus of Norwegian School English

Angela Hasselgren
University of Bergen

The English Department at the University of Bergen, home of the COLT Corpus, is in the process of launching another corpus of teenage English language, this time the speakers being Norwegian pupils, of 14–15 years. The corpus has been compiled as part of the government-sponsored EVA Project (Evaluation of English in Norwegian schools), with Angela Hasselgren as project leader and Anna-Brita Stenström as advisor. The processing of the EVA corpus has been carried at the Norwegian Computing Centre for the Humanities, along parallel lines to that of the COLT Corpus.

The main EVA corpus is made up of the transcripts of 62 pupils, in pairs with a tester (whose language is excluded from all statistical analyses) taking the EVA 8th grade oral test. This test consists of three picture-based tasks, one involving describing, narrating and discussing, one involving giving instructions and checking for understanding, and the third consisting of a semi role play, with one role fixed (read).

In addition to the main spoken corpus, which will finally consist of around 35 000 words, a smaller control corpus has been compiled, where the speakers are 16 British teenagers carrying out the same tasks.

The Norwegian pupils have each been given a global grade on their test performance. This means that pupils' language can be analysed with respect to the evaluated ability of the speaker as well as the actual task being performed.

The control corpus is valuable in providing a yardstick against which aspects of pupils language can be assessed. Furthermore, the format of the EVA query form being identical to that of COLT makes it possible for comparisons to be made between the EVA and COLT corpora, which is valuable in the further contrastive study of Norwegian pupils' English and that of native speakers. Moreover, a contrast between the British pupils' language in the EVA and the Colt corpus will offer insight on the extent to which the EVA 'test language' can be said to resemble 'real life' or authentic teenagespeak.

The first of the EVA spoken corpora, that of Norwegian pupils can currently be accessed at http://129.177.24.28/cgi-dos/eva.cmd. The control corpus is expected to be available within a few weeks. A written corpus from a similar group of Norwegian pupils, taking the EVA writing test, is expected to be launched in the course of 1997.