

Abstract

This paper describes the stages involved in implementing a corpus of spoken Irish. This pilot project (consisting of approximately 140K words of transcribed data) implements part of the design of a larger corpus of spoken Irish which it is hoped will contain approximately 2 million words when complete. It is hoped that such a corpus will provide material for linguistic research, lexicography, the teaching of Irish and for development of language technology for the Irish language.