Abs-35
Nilanjana Banerji (Educational and Children's Division, Oxford University Press), Vineeta Gupta (Educational and Children's Division, Oxford University Press), Adam Kilgarriff (Lexical Computing Ltd.), David Tugwell (Lexical Computing Ltd.), and Kate Wild (Lexical Computing Ltd.)
Oxford Children's Corpus: a corpus of writing for children
Language written for children is a text type of great interest. It needs to be lively and engaging, with simple vocabulary and structures, and is central to learning to read. We have created the Oxford Children's Corpus, a corpus of 28 million words, one third as published by OUP and two thirds from websites for children such as BBC Schools, History on the Net, Classic Readers, Nickelodeon, Neopets, Mrs. Mad, children's parts of the National Health Service, and British Museum sites. The material is fiction, textbooks, and informational and magazine material, as targeted at 7- to 14-year-olds. It is largely contemporary, though we have also included classics such as The Wind in the Willows as they are widely-read and influential. We have thoroughly cleaned, lemmatised and POS-tagged the input and loaded it into the Sketch Engine, a leading corpus tool, where it is available for research.