

Corpus File Definitions

Matthew J. Abbott

June 30, 2015

corpus.txt is the input file for E-Z Reader simulations (you can download the latest version of E-Z Reader here: <http://www.erikdreichle.com/ezreader.html>). Column definitions are as follows.

C1: Word form frequency (drawn from Francis & Kucera, 1982)

C2: Word index (first word is not considered, so always begins at 2)

C3: Cloze probability (note that cloze values may only be collected for a target word)

C4: Word string