

VOA Special English Transcripts Corpus
By Ge Xiaoshuai

Corpus Size: 2,978,205 tokens in 3,652 files.

Data Source: Data are collected from <http://gandalf.ddo.jp> . This website provides one piece of VOA special English news report each day with audio files.

Sampling Frame: We collected all the available data from the past ten years(2006.4.5-2016.6.6).

Data Size: 8.86 MB.