

Corpus specifications LnSC

Name of corpus	Late-night Show Corpus (subcorpus of the Erlangen Student Corpus Collection)
Abbreviation	LnSC
Brief description of corpus	Transcribed excerpts from late-night shows
Name of compiler	Luise Neumüller
Type of texts (e.g. genre, language)	Late-night show conversations, presentations; American English, British English, Australian English, South African English
Corpus size (in orthographic words)	8965
Number of texts/speakers	8 texts (3x Jon Stewart, 3x John Oliver, 2x Trevor Noah)
Length range of texts	5.21 min - 9.42 min
Collection period/date of texts	2011-2016
Release date/version number	11/2019
Source:	YouTube clips of <i>The Daily Show</i> (can be found on the official Comedy Central YouTube Channel)
Collection method	<ul style="list-style-type: none"> • The corpus consists of 8 transcripts from popular US-American late-night shows, which focus on the topics: guns, gun control, incidents in connection with guns and reactions to these • The corpus focuses on the linguistic complexity of the late-night show hosts. • Show: <i>The Daily Show</i>
Compilation and annotation conventions (e.g. tokenisation, lemmatisation, tagging, syntactic parsing)	<p>Based on Hutchby, I. & R. Wooffitt (2008). <i>Conversation Analysis</i>. 2nd edn. Cambridge: Polity Press. (pp. 73-87)</p> <p>(0.5) the number in brackets indicates a time gap in tenths of a second (-) very short gap = ‘latching’ between utterances or continuation of a turn across intervening lines of transcript [] square brackets between adjacent lines of concurrent speech indicate the onset and end of overlapping talk [[]] simultaneous start of speakers’ turns</p>
Sociolinguistic description of speakers (age, sex, native language, education etc.)	Jon Stewart, 49 – 53, male, American John Oliver, 36, male, British Trevor Noah, 32, male, South African Various other speakers of different ages and backgrounds
Research project for which corpus was compiled	<i>Does success of a late-night show depend on linguistic complexity of speech? A comparison of The Daily Show’s hosts Jon Stewart, John Oliver and Trevor Noah.</i> Coursework in the Hauptseminar <i>Linguistic complexity</i> . Winter semester 2015/16. FAU (Friedrich-Alexander-Universität Erlangen-Nürnberg)
Name of supervisor	Christina Sanchez-Stockhammer
Contact person/contact details	Christina Sanchez-Stockhammer christina.sanchez@lmu.de
Project homepage	http://www.erlanger-linguistik-online.uni-erlangen.de/projekte/escc.shtml
Disclaimer	You are free to use the corpora provided here for academic research. Please inform the authors of the corpora about any research projects using their material.
Citation	Neumüller, Luise. 2016. Late-night Show Corpus (LnSC). <i>Erlangen Student Corpus Collection (ESCC)</i> . http://www.erlanger-linguistik-online.uni-erlangen.de/projekte/escc.shtml