

About the Makerere Nkore-Kiga Corpus

The Makerere Nkore-Kiga corpus is one of the products from the project *Towards a corpus-based approach for describing Ugandan languages*, funded by the Andrew W Mellon Foundation through the College of Humanities and Social Sciences' Competitive Mentorship-Oriented Research Grant. Below is a quick summary of information about this corpus:

Tokens (running words): 730,039

Types (distinct words): 98,906

Files: 355

Organisation

The corpus is organized into 16 different genres/topics: Agricultural texts; Cultural texts; Environment texts; Financial texts; General development texts; Governance texts; Health texts; Historical texts; Instructional texts; Legal texts; Newspapers; Novels; Political texts; Radio News texts; Religious texts; and Social relations texts). Information about the number of files and tokens per genre is shown in the table below:

SN	Genre	#Files	Tokens	%
1	Agriculture	30	52,475	7.19
2	Culture	49	18,132	2.48
3	Environment	4	20,865	2.86
4	Finance	30	44,712	6.12
5	General Development	4	24,710	3.38
6	Governance	1	940	0.13
7	Health	36	106,830	14.63
8	History	3	16,574	2.27
9	Instruction	3	2,629	0.36
10	Law	15	118,349	16.21
11	Newspapers	120	21,692	2.97
12	Novels	19	71,639	9.81
13	Politics	8	6,960	0.95
14	Radio News	2	6,169	0.85
15	Religion	6	195,268	26.75

16	Social Relations	25	22,095	3.03
	TOTAL	355	730,039	100.00

Sources of corpus material: Internet; Scans; Translation; Transcription; Electronic Transfers

Access

Each of the 16 zipped genre folder contains a number of text files pertaining to the given genre/topic. Some files are given open access while others can only be available on request. For more information on access, please contact the corpus developer on kawalyad@chuss.mak.ac.ug

Acknowledgments:

CHUSS Grants Management Office

Assoc. Prof. Susan Kiguli, Makerere University, Uganda

Prof. Dr. Gilles-Maurice de Schryver, Ghent University, Belgium

Mr. Simon Rutahigwa