

EXHIBIT A: DESCRIPTION OF THE DATASET

AP17-OL3 Multilingual Database		Training & Development Set			Testing Set		
Language	Channel	No. of Speakers	Total Utterances	Hours	No. of Speakers	Total Utterances	Hours
Kazak	Mobile	86	4200	10	86	1800	4
Tibetan	Mobile	34	11100	12	34	1800	2
Uyghur	Mobile	353	5800	14	353	1800	4

Original scripts are available.

Lexica are available.

The database is part of M2ASR NSFC project.