

# Kurmanji Speech Corpus

The project I have in mind is a 3 phase corpus of data recorded from all Kurmanji dialects from all the geographical places that Kurmanji is being spoken. Phase I: spontaneous conversation corpus, Phase II: read corpus, Phase III: sample sentences read by speakers. The ideal goal is to record and collect all the phases; however this task is rather far-fetched and takes a lot of time. Therefore the basic phase is phase I: spontaneous conversations.

Right now I have almost covered most of the Kurmanji speaking regions of Iran and in 2-3 weeks I will be finished recording major dialects (north of Guilan province, northwest of Iran, and Quchan (in Khorasan Province)). The ideal data from Turkey would be recording several conversations from major Kurmanji dialects (1): Northwestern Kurmanji (*Meraş, Meletî, Sêwaz* provinces), Southwestern (Gaziantep, *Semsûr*), Northern Kurmanji (*Agirî, Erzerom, Mûş*), and Anatolian Kurmanji (spoken widely in Anatolia Region). Though all the dialects cannot be recorded thus feel free to alter the list.

The basic structure of the corpus is similar to Santa Barbara corpus of English, However we are including more details on profiling speakers e.g. age sex and place of living. For more information you can visit (2).

I have talked to Linguistic Data Consortium of the University of Pennsylvania for being the distributor of the corpus (for more info on distributing format for LDC visit (3)). We will approximately finish recording the corpus in the early 2016 it will be published in the second quarter of 2016. We're doing our best to provide the transcription at the time of publishing however if we were unable to achieve this we will provide the transcriptions as an update to the corpus.

(1) a: Öpengin, Ergin; Haig, Geoffrey (2014), "Regional variation in Kurmanji: A preliminary classification of dialects", *Kurdish Studies* 2, ISSN 2051-4883

(1) b [https://en.wikipedia.org/wiki/Northern\\_Kurdish](https://en.wikipedia.org/wiki/Northern_Kurdish)

(2) <http://www.linguistics.ucsb.edu/research/santa-barbara-corpus>

(3) <https://www ldc.upenn.edu/data-management/providing>