A speech recognizer for subtitling Frisian/Dutch council meetings

Henk van den Heuvel, Martijn Bentum, Louis ten Bosch, Simone Wills
CLST, Radboud University, Nijmegen

Late 2020, the Fryske Akademy was granted a project to develop a subtitling service for council meetings of Frisian municipalities. The project was financed by the Province of Fryslân, the "Wetterskip" and a number of Frisian municipalities. An existing speech recognizer for Frisian named *FAME!*, was repurposed for a new application domain: council meetings (*FAME!* was trained and tested on radio broadcasts only). The council meeting domain is difficult for speech recognition because of the acoustic background noise, speaker overlap and the jargon language typically used in council meetings.

To train the new recognizer, we used the radio broadcast materials utilized for the *FAME!* recognizer and in addition, newly created manually transcribed audio recordings of council meetings from several Frisian municipalities. The council meeting audio recordings consist of approximately 49 hours of speech, with 26 hours of Frisian speech and 23 hours of Dutch speech. Further, we obtained texts in the domain of council meetings, namely council meeting minutes and council policy documents containing approximately 11 million words; 1.1 million Frisian words and 9.9 million Dutch words. We describe the methods used to train the new recognizer, report the observed word error rates, and perform an error analysis on remaining errors.