In this paper we exploit linguistic knowledge to aid in automatic dialect identification in Spanish. Segments of extemporaneous Cuban and Peruvian Spanish dialect data from the Miami Corpus were analyzed and 49 linguistic features that occur with different rates in each of the two dialects identified and hand-labelled. We evaluate the expected performance of the dialect detection system based on a theoretical model and compute the systems' performance. Using a Gaussian classifier we show that a subset of the 49 originally-identified features obtains nearly perfect performance for discriminating between the two dialects. We compare these results with those from an automatic recognition system (PRLM-P). We then test this system in the limited domain of read digits from 0 through 10 using an orthographic transcription and hand-marked data for phone extraction and alignment. Initial experiments on phone-level segments show that phone duration and energy computations prove discriminatory for dialect discrimination.

Abstract

In this paper we exploit linguistic knowledge to aid in automatic dialect identification in Spanish. Segments of extemporaneous Cuban and Peruvian Spanish dialect data from the Miami Corpus were analyzed and 49 linguistic features that occur with different rates in each of the two dialects identified and hand-labelled. We evaluate the expected performance of the dialect detection system based on a theoretical model and compute the systems' performance. Using a Gaussian classifier we show that a subset of the 49 originally-identified features obtains nearly perfect performance for discriminating between the two dialects. We compare these results with those from an automatic recognition system (PRLM-P). We then test this system in the limited domain of read digits from 0 through 10 using an orthographic transcription and hand-marked data for phone extraction and alignment. Initial experiments on phone-level segments show that phone duration and energy computations prove discriminatory for dialect discrimination.