

Learning from 26 Languages: Program Management and Science in the Babel Program

Mary Harper

Incisive Analysis Office

Intelligence Advanced Research Projects Activity

Office of the Director of National Intelligence

USA

`mary.harper@iarpa.gov`

Invited Speaker Abstract

This presentation will illustrate how program management and science can cooperate to increase our understanding of human languages and algorithms for processing them. In this presentation, I will use the IARPA Babel program as an example. The goal of the Babel Program is to rapidly develop speech recognition capability for keyword search in new languages, working with speech recorded in a variety of conditions and with limited amounts of transcription. The speech data is recorded in native countries and contains variability in speaker demographics and recording conditions. The Program will ultimately address a broad set of languages with a variety of phonotactic, phonological, tonal, morphological, and syntactic characteristics. I will discuss the data resources collected to support the research, the challenges that performers have faced when working with a variety of languages collected in realistic environments, the lessons learned, and future directions.