Discriminative training approaches aim to maximise:

1. Background
   - Language Identification (LID) is key to spoken dialogue systems that must function in multilingual environments
   - South Africa officially recognises 11 languages
   - Most citizens are fluent in more than one language
   - We focus on four languages that together account for 63% of mother-tongue speakers: Afrikaans, South African English, Xhosa and Zulu

2. Databases
   - Telephone speech annotated both orthographically and phonetically
   - Mix of spontaneous and read speech

3. System Architecture
   - Speech as unknown language
   - Transcription and language classification
   - VITERBI RECOGNISER
   - MULTILINGUAL LANGUAGE MODEL
   - MULTILINGUAL ACOUSTIC MODEL

4. Discriminative Training
   - Discriminative training approaches aim to maximise:
     \[ p_d(X_i | L_i) \approx p_d(X_i | W_i) \]
   - Discriminative training has narrowed gap between short and long utterances
   - HD2 improves LID for both short and long utterances
   - HD2 shows LID improvement on both training and test-sets
   - ASR performance hardly affected

5. Utterance length
   - Average length is short: 2.1s for Afrikaans and English, 2.6s for Xhosa and Zulu
   - Shorter utterances: Model Longer utterances

5. Code mixing
   - English often used to cite numbers, dates and times in modern Xhosa and Zulu
   - 27.9% of Xhosa and 31.2% of Zulu test utterances contain 75% or more English words
   - LID is less accurate for code-mixed utterances
   - Discriminative training worsens LID performance for code-mixed utterances