A Tale of Two Phonologies: Problems in Armenian ASR

Samuel Chakmakjian

A Tale of Two Phonologies: Problems in Armenian ASR

INTRODUCTION

Armenian is a traditionally under-resourced language, which has seen a recent uptick in interest in the development of its tools and presence in the digital domain. Some of this recent interest has centered around the development of Automatic Speech Recognition (ASR) technologies. However, the language boasts two standard variants which diverge on several levels, the most salient of which being on the phonetic-phonological level. Despite this divergence several social factors render a unified ASR system preferable and practical. These factors include :

- a high level of mutual-intelligibility
- intense social overlap of the speech communities
- increased contact between the two; bi-variant households and speakers

THE PROBLEM

In terms of ASR, the most problematic feature of the divergence between Standard Western Armenian (SWA) and Standard Eastern Armenian (SEA), is that of the phonologies, and more specifically, the voicing systems.

SWA (Western Armenian)

- Traditionally: Anatolia, (post-)Ottoman zone
- Currently: Diasporan communities in Middle East, Europe and North America

- 30 phonemes : 24 consonants, 6 vowels
- **Two-way voicing system**
- Voiced; voiceless aspirated

	Bil.	L-d.	Alv.	P.-Al.	Pal.	Vel.	Uv.	Glot.
Nas.	m		n					
Plos.	pʰ b		tʰ d			kʰ g		
Affr.			ts dz	tʃ dʒ				
Fric.		f v	s z	ʃ ʒ			χ ʁ	h
Appr.			l		j			
Tap			ɾ					
Trill								

SEA (Eastern Armenian)

- Traditionally: Caucasus & Iran
- Currently: Republic of Armenia; Iran, post-Soviet countries, Europe and North America

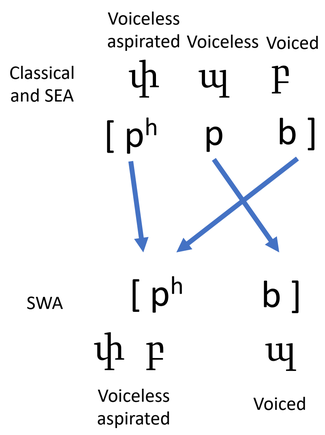
- 36 phonemes : 30 consonants, 6 vowels
- **Three-way voicing system**
- Voiced; voiceless; voiceless aspirated

	Bil.	L-d.	Alv.	P.-Al.	P	Vel.	Uv.	G
Nas.	m		n					
Plos.	pʰ p b		tʰ t d			kʰ k g		
Affr.			tsʰ ts dz	tʃʰ tʃ dʒ				
Fric.		f v	s z	ʃ ʒ			χ ʁ	h
Appr.			l		j			
Tap			ɾ					
Trill			r					

Armenian's orthography (in both variants) maintains a representation of three graphemes for each of these voicing sequences, making rule-based speech synthesis of either pronunciation feasible from the same text. However, producing text from speech poses a challenge when some acoustically identical inputs correspond to different graphemes in the two variants.

SEA	SWA
Voiced	Voiceless aspirated
Voiceless	Voiced
Voiceless aspirated	Voiceless aspirated

item	SEA	SWA	translation
⟨ բառ ⟩	[**b**ɑr]	[**p**ʰɑɾ]	'word'
⟨ պար ⟩	[**p**ɑɾ]	[**b**ɑɾ]	'dance'
⟨ փառ ⟩	[**p**ʰɑr]	[**p**ʰɑɾ]	'placenta'



THE DATA

Our work to construct an ASR model for Armenian is conducted in the framework of the ANR-funded DALiH (*Digitizing Armenian Linguistic Heritage*) project, within which we expect to take advantage of two major transcribed audio corpora.

Rerooted Archives

- Western Armenian (speakers from Syria)
- corpus of interviews, monologues
- 120 hours of audio data
- 82 hours transcribed, and aligned by sentence

Eastern Armenian National Corpus (EANC)

- Eastern Armenian
- corpus of spontaneous, public and task-oriented discourse; online communications
- 774 total transcribed (audio) documents
- 3.5 million tokens

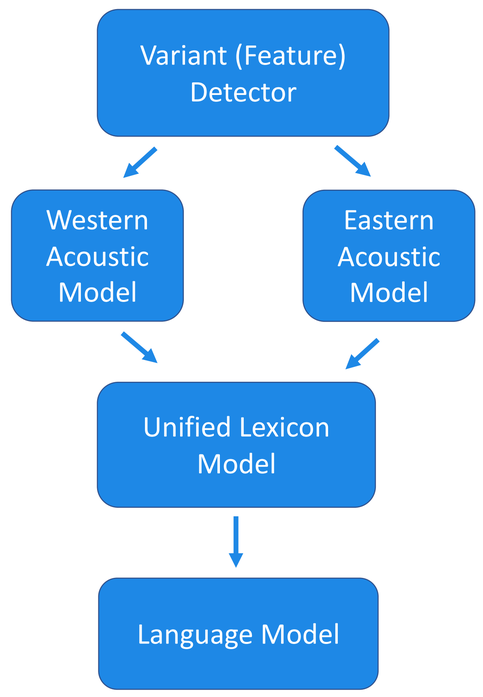
SOLUTIONS

To tackle this problem, we will first construct a traditional hybrid ASR model, with the eventual goal of comparing the efficacy of the system to an End-to-end (E2E) model. In these tests, we will use the manually transcribed and aligned data from the EANC and Rerooted corpora to determine the level of accuracy yielded by the two methods.

Below, we propose two pathways for processing audio data from the two dialects in one (hybrid) system.

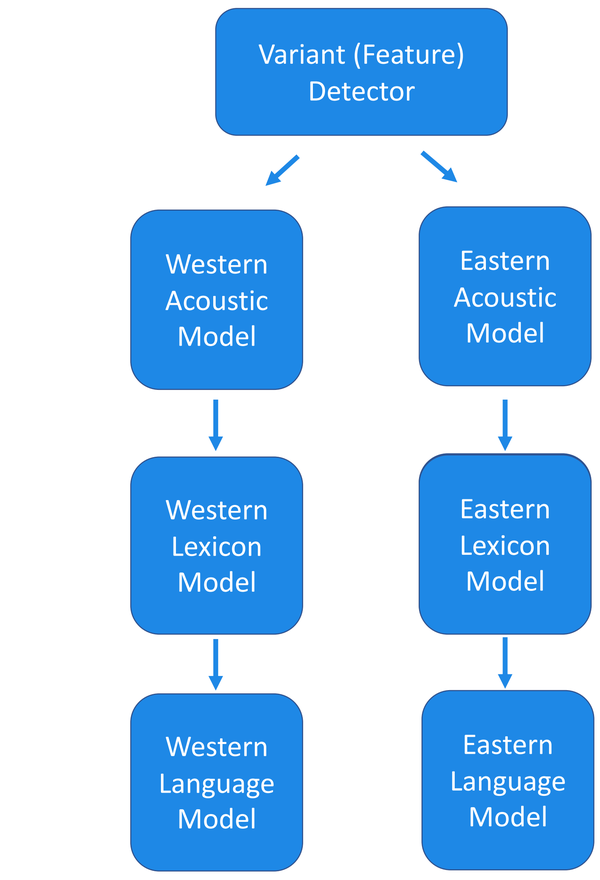
OPTION 1

Two separate acoustic models, which feed into a common lexicon model.



OPTION 2

Two completely independent paths for the two variants.



Samuel CHAKMAKJIAN
Սամուէլ Չաքմաքճեան
Doctoral student, Inalco SeDyL & ERTIM