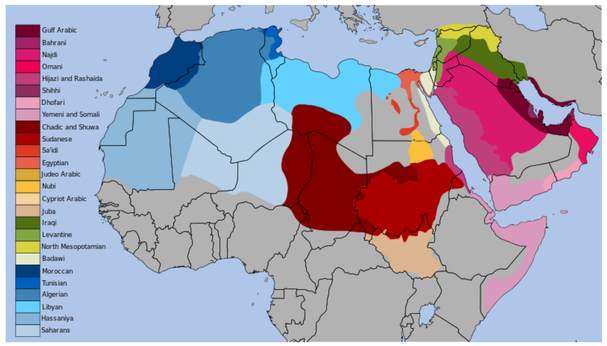
Lexical Modeling for Arabic ASR: A Systematic Approach

Tuka Al Hanai, James Glass
MIT Computer Science and Artificial Intelligence Lab, Cambridge MA, USA

1. Overview



Challenges of Arabic Lexical Modeling

- Deep Orthography
  - No diacritics – 30% of characters 'missing'
  - Ambiguous word to phoneme mapping

wktb AlkAtb	وكتب الكاتب
wakataba AlkAtibu	وَكَتَبَ الْكَاتِبُ

Research Question

How does lexical modeling with
- Diacritics
- Pronunciation rules
- Acoustics

affect ASR performance?

Literature presents

- Mixed adoption of diacritics
- Little investigation of pronunciation rules
- Stochastic lexicon in Arabic relies on an existing database of pronunciations

ASR System

- 70 hours MSA broadcast news LDC GALE
- Triphone HMM-GMM acoustic models (Kaldi)
- Trigram LM with KN smoothing (SRILM)
- Transcripts diacritized using MADA Toolkit

2. Experiments

I. Diacritics

- Diacritics categorized with respect to behavior
- Grapheme can have multiple phonetic mappings
- MADA produces
  - 5.42 pronunciations per word in lexicon
  - 6.73 in text.

Category	Short Vowels			
Diacritic	a	u	i	o
Arpabet	/ae/	/uh/	/ih/	null
Example	*kataboti* / k ae t ae b t ih /			

Category	Geminates
Diacritic	~ (tilde)
Example	*kataba* / k ae t ae b ae / *kat~aba* / k ae t t ae b ae /

Category	Nunnations		
Diacritic	F	K	N
Arpabet	/ae n/	/uh n/	/ih n/
Example	*kitAban* / k ih t ae: b ih n /		

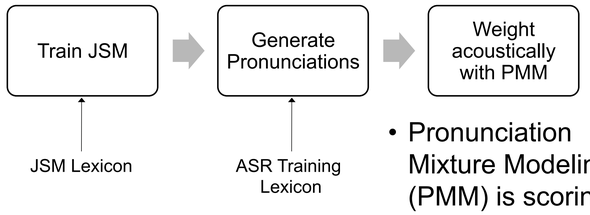
II. Pronunciation Rules

- Evaluating lexicons with 3 descriptions of pronunciation rules from the Arabic literature
  - Rules I – *Ahmed* [3]
  - Rules II – *El-Imam* [4]
  - Rules III – *Biadsy et al.* [5]
- Rules I and II designed for TTS system, Rules III derived and evaluated for ASR.

Modeling	Rules I	Rules II	Rules III
Glottal stops ({ ', }, &, <, >} → q)	✔		✔
Short vowels	✔	✔	✔
Co-articulation of definite article *al* (alshams → / ae sh ae m s /)	✔	✔	✔
Nunnations	✔	✔	✔
Diphthongs	✔	✔	✔
Geminates	✔		
Word ending *p* (p → t | ε)	✔		✔
Phones in pharyngealized contexts	✔	✔	
Phones in emphatic contexts	✔	✔	
Phones in aspirated contexts	✔		
Retroflexed Vowels	✔		
Case endings			✔
# phones	135	63	34

III. Acoustic Based Pronunciation Modeling

- Joint Sequence Models (JSM) model relationship between grapheme and phonetic units



- Pronunciation Mixture Modeling (PMM) is scoring pronunciations over N-best list of hypotheses

Alshms / A sh ae m s /	=	A A	l ε	sh sh	ε ae	m m	s s

- Improvement when using a lexicon generated statistically and weighted acoustically
- Limitation with the source of JSM training

3. Results

Lexicon	Freq. (%)	PPW*	WER (%)	Sig. *p* <
No Diacritics	-	1	25.1	-
Short Vowels	25	1.25	24.1	0.007
No Nunnations	1	1.28	23.9	0.001
No Geminates	3	1.25	23.2	0.001
All Diacritics	29	1.28	23.4	0.001

*PPW = Pronunciations Per Word

- Short vowels & nunnations provide an almost equal gain
- Geminates hurt, likely due to data sparsity

Lexicon	# phones	PPW*	WER (%)	Sig. *p* <
No Diacritics	36	1	25.1	-
Diacritics	72	1.28	23.4	0.001
Rules I	135	1.27	24.0	0.004
Rules II	63	1.27	23.3	0.001
Rules III	34	1.85	22.7	0.001

*PPW = Pronunciations Per Word

- Simple rules are most effective
- Focus on
  - Capturing coarticulation (*al* & case endings)
  - Eliminating modeling redundant grapheme units (glottal stops)

Lexicon	JSM Lex Vocab (PPW*)	PMM Lex PPW*	WER (%)	Sig. *p* <
Baseline	-	1.85	22.7	-
GALE + MADA – all prons	-	1.57	22.4	0.254
GALE + MADA – top prons	61K (1.28)	2.05	22.3	0.407
GALE + MADA – all prons	61K (5.42)	2.02	22.0	0.026
Nemlar + manual diac	39K (1.56)	2.05	22.5	0.159
Nemlar + MADA – top prons	36K (1.29)	2.02	22.2	0.689
Quran + manual diac	15K (1.18)	1.93	23.9	0.001

PPW = Pronunciations Per Word.
M-gram = 5, L = R = 1, K = 50 candidates, N-best = 100, T = 0.01.

4. Conclusion

Acoustics allow for better pronunciation modeling than a manual lexicon.

5. Future Work

- Derive diacritics from acoustics
  - Allows skipping of NLP preprocessing
  - Can aid in understanding speech
- Apply to lexical modeling of Arabic dialects when lexicons and NLP toolkits are lacking
- Exploring these techniques with DNN