Challenges and Techniques for Dialectal Arabic Speech Recognition and Machine Translation

Mohamed Elmahdy, Mark Hasegawa-Johnson, Eiman Mustafawi, Rehab Duwairi, Wolfgang Minker

Nov. 21, 20011

Qatar University
University of Illinois
Ulm University

Arabic Language

- Largest still living Semitic language
- 250+ million native speakers

Arabic

Formal

- Modern Standard Arabic (MSA)
- Standardized
- A lot of ASR and MT research
- Not used in everyday life

Dialectal

- Used in everyday life
- Not standardized (mainly spoken)
- Many different dialects
- Very few ASR and MT research

Significant differences between MSA and Dialectal Arabic

➢ *Considered as completely different languages*

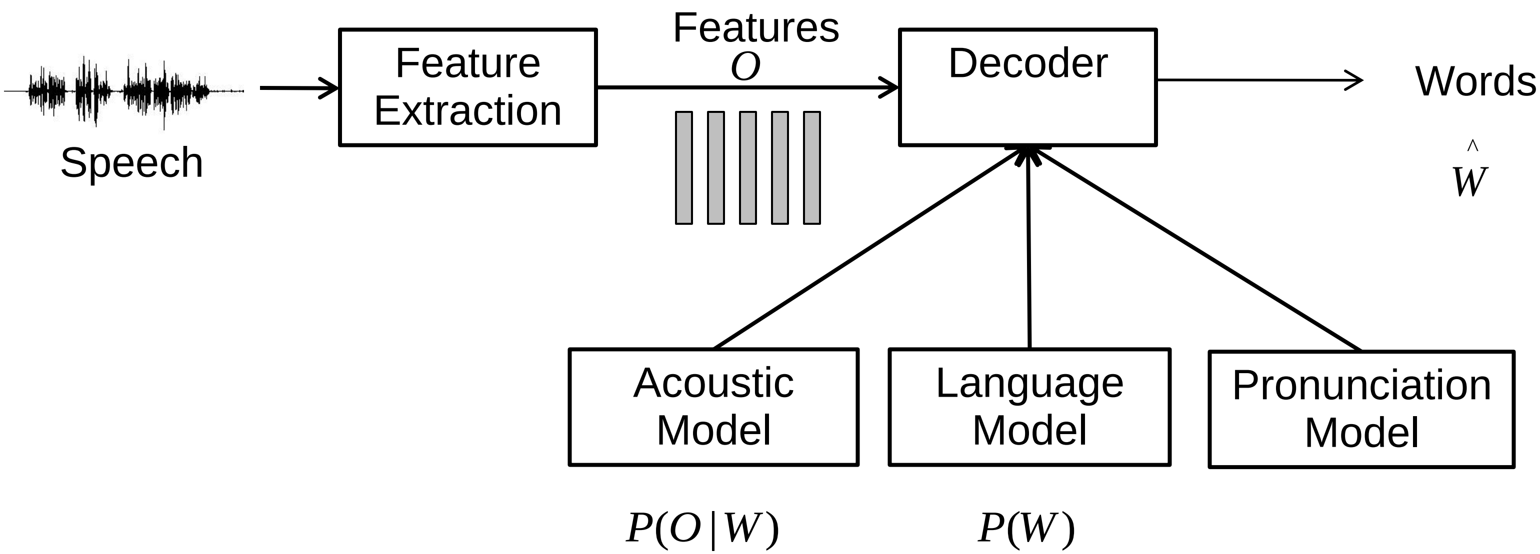
MSA Versus Dialectal Arabic

- Let’s have Egyptian Colloquial Arabic (ECA) as a typical Arabic dialect
- Phonological
  → /t/, /s/ in ECA instead of /T/ in MSA
  e.g. /tala:tah/ (three) in ECA versus /Tala:Tah/ in MSA
- Lexical
  → /t‘ArAbE:zA/ (table) in ECA versus /t‘awila/ in MSA
- Syntactic
  → SVO in ECA versus VSO in MSA

Automatic Speech Recognition

- High level diagram for a state-of-the-art ASR system

$$\hat{W} = \underset{W \in L}{\arg\max}\, P(O|W)P(W)$$

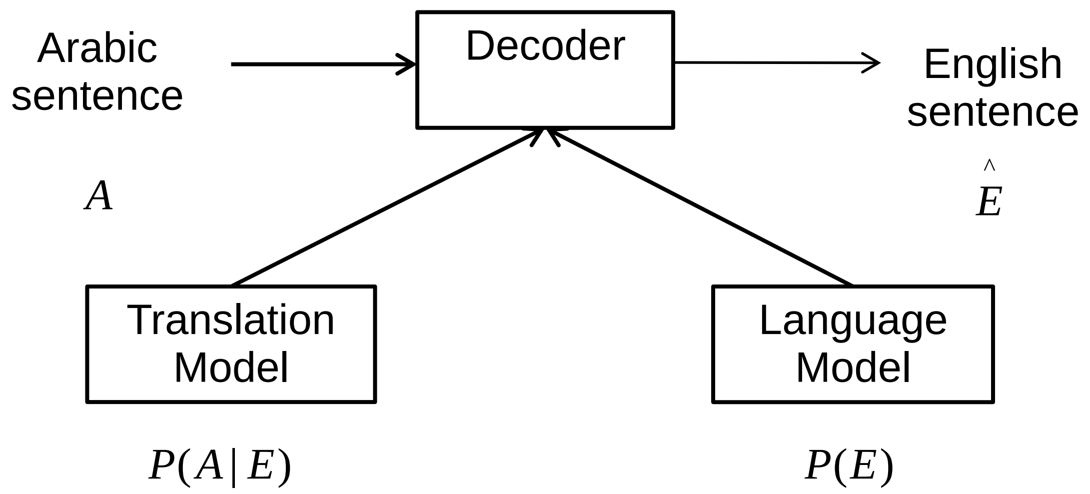


For dialectal Arabic, sparse and low quality corpora are available

Statistical Machine Translation

- High level diagram for a SMT system

$$\hat{E} = \arg\max_{E \in English} P(A \mid E) P(E)$$



Large parallel corpora are required
For dialectal Arabic, parallel corpora are not available

Objectives

- ASR and MT for dialectal Arabic where little data exists
- To benefit from existing MSA speech data to improve dialectal Arabic ASR and MT
- Ultimate goal “Speech-to-text MT” for dialectal Arabic

Outline

- Introduction
- Approaches
- Experiments and results
- Conclusions and future directions

Proposed Approaches for Dialectal Arabic ASR

- **Phonemic acoustic modeling**
  → Dialectal speech data where phonetic transcription is available
- **Graphemic acoustic modeling**
- **Unsupervised acoustic modeling**
- **Arabic Chat Alphabet-based acoustic modeling**

Phemic Cross-Lingual Acoustic Modeling

- Benefit from existing large MSA speech corpora

- Assumptions:
  - → MSA is always a 2nd language for any Arabic speaker
  - → Large amount of MSA speech data (large number of speakers) implicitly cover all the acoustic features of the different Arabic dialects

- Approach:
  - → Train an acoustic model using a large amount of MSA speech data
  - → Adaptation of the MSA acoustic models with a little amount of dialectal speech data

Phonemic Cross-Lingual Acoustic Modeling (cont.)

- State-of-the-art AM adaptation techniques include:
  - → Maximum Likelihood Linear Regression (MLLR)

$$\Phi_{MLLR} = A\Phi + b$$

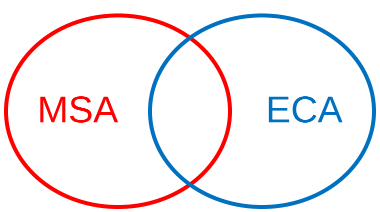
  - → Maximum A-Posteriori (MAP)

$$\Phi_{MAP} = \arg\max_{\Phi} P(O \mid \Phi) P(\Phi)$$

  - ➢ Requirement: adaptation data and the AM have to share the same language and phoneme set

- Egyptian Colloquial Arabic (ECA) is chosen as a typical dialect

- INITIALLY: MSA and ECA do not share the same phoneme inventory

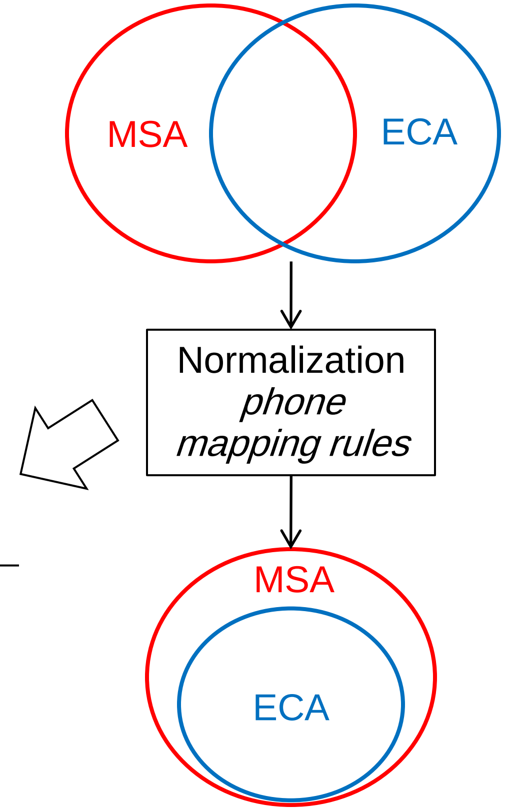


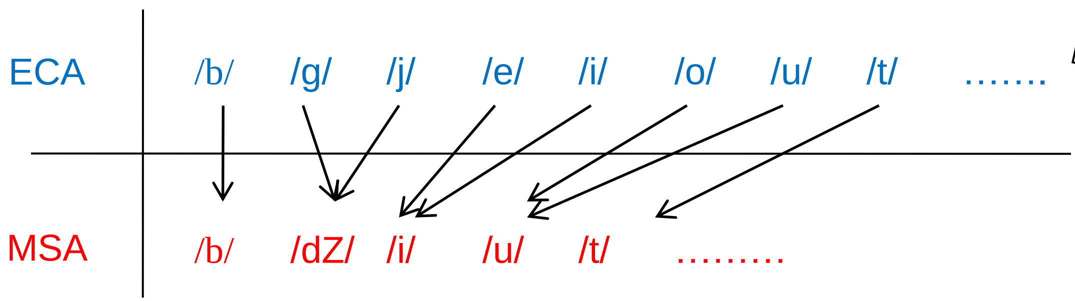
→ *Acoustic model adaptation is not possible*

Phonemic Cross-Lingual Acoustic Modeling (cont.)

- SOLUTION: Phoneme sets normalization
  → AM adaptation is possible

- Phoneme sets normalization
  → Several phone mapping rules are applied
  → Map ECA phonemes to their origins in MSA (even if they are acoustically different)



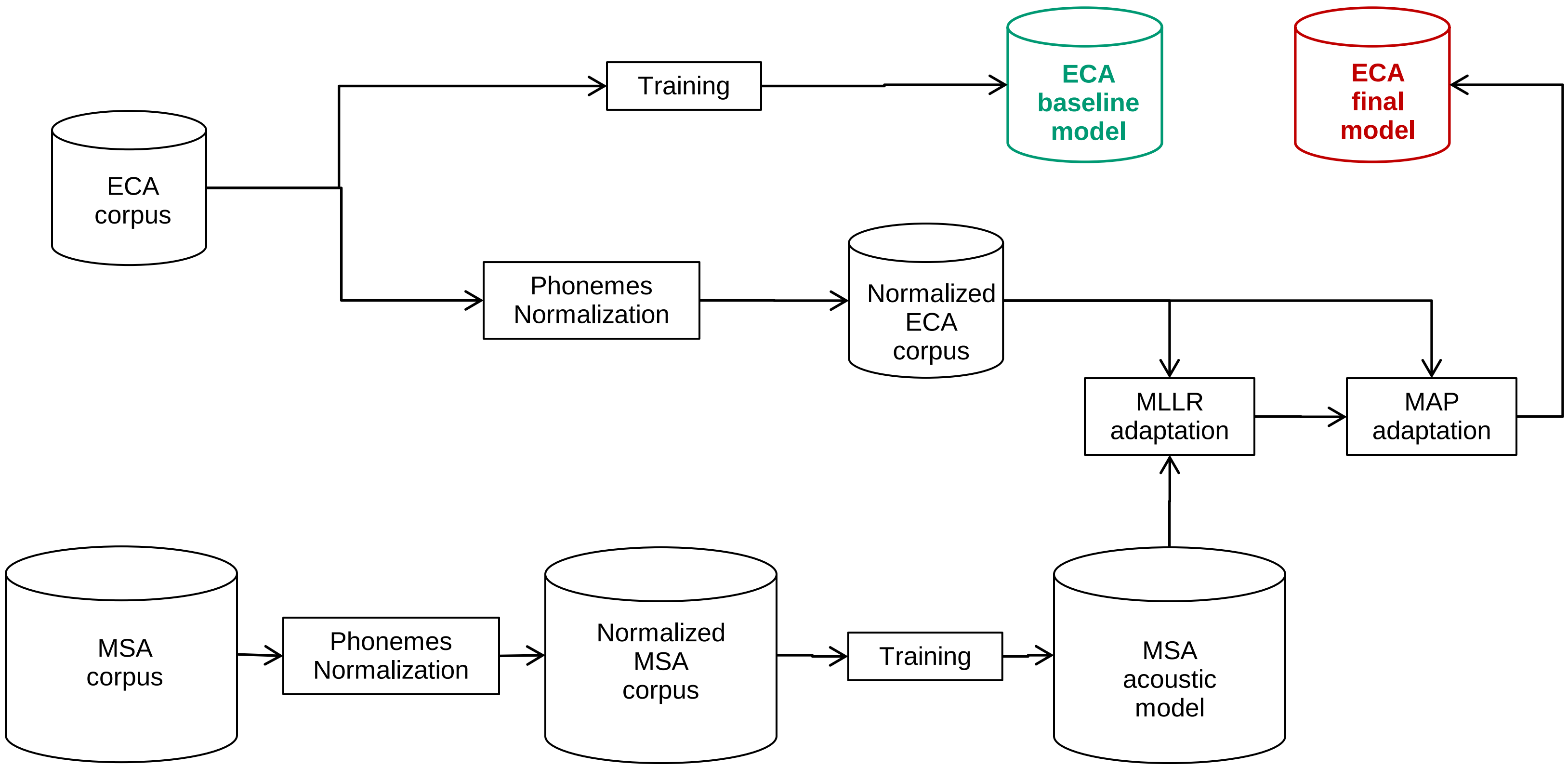


جزر
(carrot)

/g/ /A/ /z/ /A/ /r/ ⟶ /dZ/ /a/ /z/ /a/ /r/

Phonemic Cross-Lingual Acoustic Modeling (cont.)

- Block diagram for the proposed approach
- The adapted ECA AM is evaluated against the ECA baseline AM



Proposed Approaches for Dialectal Arabic ASR

- **Phonemic acoustic modeling**
  - → Dialectal speech data where phonetic transcription is available
- **Graphemic acoustic modeling**
  - → Phonetic transcription is not possible/difficult
  - → Short vowels are missing
  - → Phonetic transcription is approximated to be word letters
- **Unsupervised acoustic modeling**
  - → Transcriptions are not available at all
  - → Dialectal speech was automatically transcribed using a MSA model
- **Arabic Chat Alphabet-based acoustic modeling**
  - → Latin letters are used instead of Arabic ones
  - → Include short vowels that are missing in traditional Arabic orthography

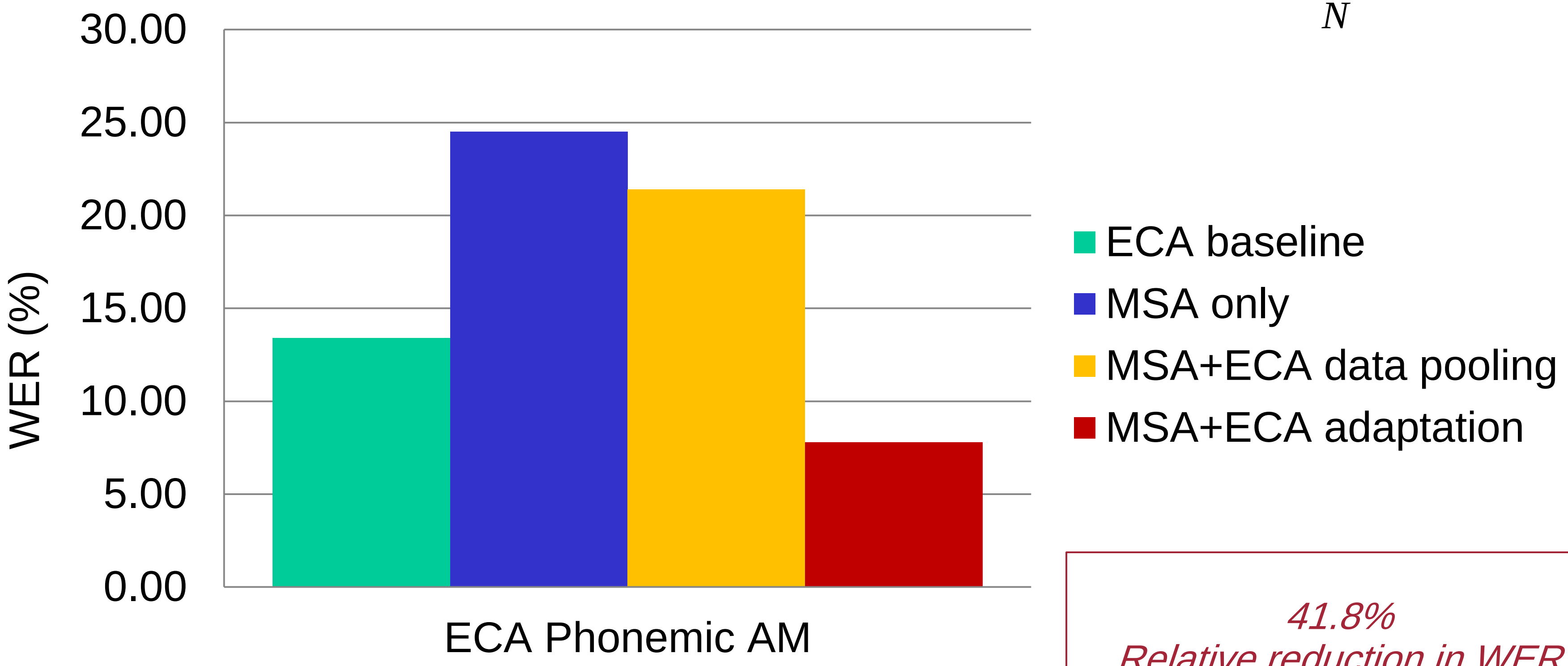
Outline

- Introduction
- Approaches
- Experiments and results
- Conclusions

Phonemic Cross-Lingual Adaptation Results

- ECA corpus:
  - → 65% for training/adaptation
  - → 35% for testing
- Word Error Rate (WER)

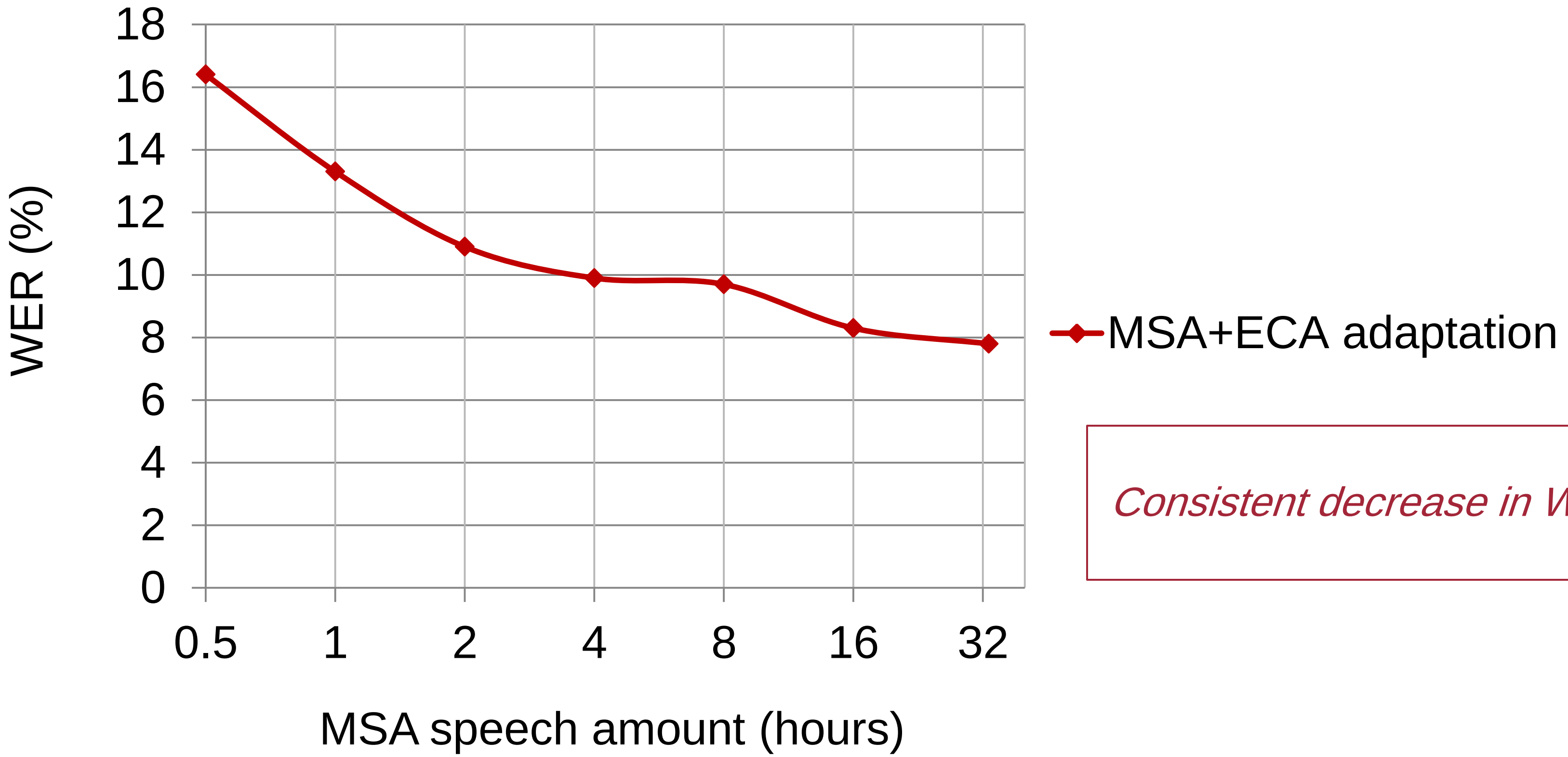
$$WER = \frac{Sub + Ins + Del}{N}$$



41.8%
Relative reduction in WER

Effect of MSA Speech Data Amount

- Varying the amount of MSA speech data
- Effect on phonemic cross-lingual adaptation



Consistent decrease in WER

Outline

- Introduction
- Approaches
- Experiments and results
- Conclusions

Conclusions and Future Directions

- **Conclusions**
  - → Problems in ASR and MT for dialectal Arabic
  - → Cross-lingual acoustic modeling for dialectal Arabic ASR
  - → Improvements are observed in both phonemic and graphemic modeling
  - → Consistent reduction in WER by adding more MSA data

- **Future directions**
  - → Data collection (a focus is placed on the Qatari dialect)
  - → Extension to all the Arabic dialects
  - → Dialectal Arabic MT and LM

Thank you for your attention