Development of the Arabic Loria Automatic Speech Recognition system (ALASR) and its evaluation for Algerian dialect

Mohamed Amine Menacer, Odile Mella, Dominique Fohr, Denis Jouvet, David Langlois, Kamel Smaïli

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2017) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

3rd International Conference on Arabic Computational Linguistics, ACLing 2017, 5–6 November 2017, Dubai, United Arab Emirates

Development of the Arabic Loria Automatic Speech Recognition system (ALASR) and its evaluation for Algerian dialect

Mohamed Amine Menacer[a], Odile Mella[a], Dominique Fohr[a], Denis Jouvet[a], David Langlois[a], Kamel Smaïli[a]

[a]*Loria, Campus Scientifique, BP 239, Vandoeuvre Lès-Nancy, 54506, France*

Abstract

This paper addresses the development of an Automatic Speech Recognition system for Modern Standard Arabic (MSA) and its extension to Algerian dialect. Algerian dialect is very different from Arabic dialects of the Middle-East, since it is highly influenced by the French language.
In this article, we start by presenting the new automatic speech recognition named ALASR (Arabic Loria Automatic Speech Recognition) system. The acoustic model of ALASR is based on a DNN approach and the language model is a classical n-gram. Several options are investigated in this paper to find the best combination of models and parameters. ALASR achieves good results for MSA in terms of WER (14.02%), but it completely collapses on an Algerian dialect data set of 70 minutes (a WER of 89%). In order to take into account the impact of the French language, on the Algerian dialect, we combine in ALASR two acoustic models, the original one (MSA) and a French one trained on ESTER corpus. This solution has been adopted because no transcribed speech data for Algerian dialect are available. This combination leads to a substantial absolute reduction of the word error of 24%.

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Arabic Computational Linguistics.

Keywords: speech recognition; Algerian dialect; language model; acoustic model; MSA.

1. Introduction

Modern Standard Arabic, generally referred as MSA (Alfus'ha in Arabic), is the variety of Arabic that was retained as the official language in all Arab countries, and as a common language between them. Standard Arabic is not acquired as a mother tongue, but rather it is learned as a second language at school and through exposure to formal broadcast programs (such as the daily news), religious practice and newspapers reading. MSA is characterized by a complex morphology and a rich vocabulary, since it is a very old language, and it can still be understood in its original form in a better way than French or English. It is an inflectional and agglutinative language; let's recall that compared to English, an Arabic word (or more rigorously a lexical entry) can sometimes correspond to a whole English sentence.

* Corresponding author. Tel.: +33-755-260-590.
E-mail address: mohamed-amine.menacer@loria.fr

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Arabic Computational Linguistics.

The grammatical system of Arabic language is based on a root and pattern structure, which uses more than 10000 roots and 900 patterns [5]. Arabic words are derived from roots by using patterns.
However, most of Arab people do not use MSA in their daily conversations, since their mother tongue is an Arabic dialect, which is derived from MSA. Furthermore, for each region, there is one or several dialects influenced by the history of the region itself. New words borrowed from English, Turkish, Spanish, Italian or French are integrated in the vocabulary of these dialects. Some of these new words are used such as in the original language, but others are altered in order to respect the morphological structure of the Arabic language. For instance, the French word *casseroles* (*pans*), which is pronounced /k a s r o l/, has been borrowed and integrated in Algerian dialect. It is written in this dialect as follows: كسرونَات, which is uttered /k a s r uː n aː t/.

Our objective in this paper is to present the development of an Automatic Speech Recognition (ASR) system for MSA (namely ALASR: Arabic Loria Automatic Speech Recognition), to evaluate its performance on MSA data and then to test it on a corpus that is not entirely MSA. This corpus is composed of Algerian dialect sentences extracted from PADIC [17].
The development of such system requires to take into account the characteristics of MSA (highly inflectional and agglutinative language), several studies have dealt with some of these issues. In [10], [11], [19] and [9], the authors proposed to use morphological segmentation to reduce the size of the lexicon. In some of these works they maintain the quality of the ASR and for others they improve the quality of their systems.

The absence of diacritics in Arabic texts is a serious issue for many applications in Natural Language Processing (NLP). For every Arabic root which is not vocalized, the ASR system has to consider all the possibilities of pronunciations or has to restore the diacritics. For example, the word قبل can be pronounced such as: قَبلَ /q a b l a/ (*before*), قَبِلَ /q a b i l a/ (*accepted*) or قُبِّلَ /q u bː i l a/ (*was kissed*) and it also has other diacritizations, which are not listed here.
It is shown in [10] that using short vowels in the training transcripts and applying a morphological decomposition on the vocabulary, improved the Word Error Rate (WER).
In speech recognition, the bulk of works proposed in the literature is intended for MSA, while only few works are dedicated to Arabic dialects. This is mainly due to the lack of Arabic dialect data [25, 6]. The efforts made, until now, to develop ASR for Arabic dialects concern those considered as close to MSA, namely Iraqi [3], Egyptian [7], Qatari [13] and Leventine [23, 12, 21]. However, there are few ASR systems for Maghrebi dialects especially those used in Algeria. The main characteristic of Algerian dialect is the code-switching, meaning a speaker alternates between MSA and the native dialect, which is influenced by French, Berber and Turkish.
In this work, we decided to test ALASR system on Algerian dialect since it is difficult to develop an entire ASR due to the lack of data necessary for training the acoustic model. To deal with this issue, some changes on the acoustic and language models have been made, in order to make ALASR system doing its best on Algerian dialect.
In the next section, a description of ALASR system is presented. Section 3 gives an overview of the differences between MSA and Algerian dialect and presents the modifications adopted in order to make ALASR usable with Algerian dialect. Experimental results are discussed in Section 3.4.

2. ALASR: Arabic Loria Automatic Speech Recognition system

A speech recognition system needs at least two components: an acoustic model and a language model. In the following, each component is described by presenting the different steps required to train each model. Training necessitates two kinds of data: acoustic and textual, which are presented in the following.

2.1. Acoustic model

The development of the acoustic model is based on Kaldi [20], which is a state-of-the-art toolkit for speech recognition based on Weighted Finite State Transducers [18]. The recipe used relies on 13-dimensional Mel-Frequency Cesptral Coefficients (MFCC) features with their first and second order temporal derivatives, which leads to 39-dimensional acoustic features. For Arabic, 35 acoustic models (28 consonants, 6 vowels and silence) are trained. The emission probabilities of the HMM models are estimated by DNN (namely DNN-HMM). The DNN-HMM are trained

by applying sMBR criterion [24] and using 40 dimensional features vector (fMLLR) [14] for speaker adaptation. The topology of the neural network is as follows: a 440-dimensional input layer (40 × 11 fMLLR vectors), 6 hidden layers composed of 2048 nodes each and a 4264-dimensional output layer, which represents the number of HMM states. The total number of weights to estimate is about 30.6 million.

2.2. Language model

In Arabic, even in newspapers, few words could be simplified especially at the beginning or at the end by replacing a specific letter by another one or by omitting the *hamza* symbol. Unfortunately, both written forms of a word may exist in a same document. This leads to share a probability over two forms that correspond to the same word. For instance, the word إستعمَال (*uses*) is the right way to write it, but people could omit the *hamza* and write it such as: استعمَال. Obviously, this is not the only case, several other points have to be treated. Some of them are specific to Arabic and others are used in the majority of natural languages. That is why, in NLP applications dedicated to Arabic, a preprocessing of the textual data is necessary before calculating the language model. For our case, the rules described in Table 1 are adopted.

Table 1. List of the preprocessing operations

Phrase	**Operation**
Email addresses - URL paths - Special characters	
Punctuation - Non-Arabic texts - Diacritics	Remove
Numbers	Conversion into words
The prefix و (*and*)	Separation from the following word, by using Farassa toolkit [1]
The prefixes ب, ف, ال, ك, ل and س	Concatenation to the following word
The stretched words	Suppression of the duplicated letters الرجاااال → الرجال
ة	Insertion of a space after this letter (this letter is always written at the end of the word)
Time	Writing it literally (15:30 → الثالثة و ثلاثون دقيقة)
Shortened forms	Replacement by the sequence of corresponding words

The operations presented in Table 1 are applied to the text corpus, which is composed by GigaWord and speech transcripts used to train the acoustic model. As this data set is unbalanced, a 4-gram language model has been developed for each part of this corpus, and they are then combined linearly. The optimal weights are determined in order to maximize the likelihood of the development corpus. Due to memory constraints, the full 4-gram language model has been pruned by minimizing the relative entropy between the full and the pruned model [22].

2.3. Vocabulary

In ASR, the vocabulary has to be representative of what should be recognized by the system. That is why, we first selected the 109K most frequent words from GigaWord and all the words that occur more than 3 times in the corpus corresponding to the speech transcripts. Afterwards, only the words, for which at least one pronunciation does exist in an external lexicon [8], are kept. In fact, this lexicon is used as a lookup table from which the pronunciations of the selected words are extracted and inserted in the pronunciation lexicon of the ALASR system. This process produces a vocabulary of 95K unique grapheme entries that correspond to 485K pronunciation variants. In other words, each grapheme entry has an average of 5 pronunciation variants.

2.4. Data

Concerning the acoustic data, 63 hours extracted from Nemlar[1] and NetDC[2] corpora have been used. They correspond to MSA broadcast news recorded from four different radio stations: Medi1, Radio Orient (France), RMC

[1] http://catalog.elra.info/product_info.php?products_id=874
[2] http://catalog.elra.info/product_info.php?products_id=13&language=fr

(Radio Monte Carlo) and RTM (Radio Television Maroc) between 2001 and 2005. The data are split randomly into three parts (Train, Dev and Test): 83% are used for training (315K words), 9% for tuning (31K words) and the rest (8%, 31K words) for evaluating the performance of the system.
Concerning the language model, a collection of textual data is required. Two corpora are used in our case: the transcripts of the acoustic training data and the Gigaword[3] Arabic corpus. The transcripts of the acoustic training data contain 315K words with a total of 34K unique words. The Gigaword corpus consists of more than one billion words, which corresponds to 4 million unique words.

2.5. Evaluation

We evaluated ALASR on the MSA test data presented in the previous section. The results, in terms of Word Error Rate (WER) are reported in Table 2. The baseline system of ALASR achieves a WER of 15.32 on the test set; it should be noted that the Out-of-Vocabulary (OOV) rate is 2.54%. In order to study the impact of the pruning on the performance of ALASR, experiments have been carried out with 3-gram and 2-gram language models (System 1 and 2 in Table 2) instead of 4-gram. Experiments show that with a smaller n-grams the results are better than with a pruned 4-gram. This is due to the fact that even if 4-grams provide longer phrases, shorter n-grams propose more phrases and then may have a better representation of the language. For instance, the number of bigrams is 5 times greater than the number of phrases of size 4 in the pruned 4-gram language model.

Table 2. WER (%) of ALASR, before and after rescoring the lattice produced by the baseline system (AM: Acoustic Model, LM: Language Model, Lex: Lexicon).

System	AM	LM	Lex	Dev WER	Test WER
Baseline	MSA	pruned 4-gram	MSA	14.65	15.32
System 1	MSA	3-gram	MSA	13.61	14.52
System 2	MSA	2-gram	MSA	13.54	14.42
Rescoring	MSA	full 4-gram	MSA	13.07	14.02

In order to improve the results, we decided to rescore the lattice produced by the pruned 4-gram model. The rescoring, in this step, consists in using a better language model since, the one adopted for producing the lattice is pruned due to the memory limitation. When the lattice is generated, there is no more constraint of memory limitation and a more precise language model may be used. That is why, the previous probabilities are replaced by those provided by a full 4-gram without any pruning. Then the engine of ALASR generates a new hypothesis of recognition. This rescoring process outperforms the baseline system by 1.3%.

3. Testing ALASR on Algerian Arabic

In the previous sections, we presented ALASR, which is an automatic speech recognition system for MSA. Recognition of Arabic dialect is an interesting challenge, since data for training are not available, especially for Algerian Arabic. People, who are not familiar with Algerian dialect, could think that the issue is not complicated, since the Arabic ASR is available, one needs just few adaptations to make it working for Algerian dialect. In the following, some descriptions about this dialect are presented and some experiments are achieved showing that ALASR completely collapses on dialectal data.
In Algeria several dialects are used in daily communication. In this work, we deal with the one used in Algiers and its periphery. The vocabulary used in Algiers dialect is influenced by MSA, French, Berber and Turkish. Concerning the words of MSA used in the dialect, some of them are pronounced as in MSA, but for others a phonological alteration is done to make them easier for the pronunciation. An example of an altered MSA word is بيت (*house*), which is pronounced /b a j t/ in MSA and /b iː t/ in Algerian dialect. As regards to the usage of new words, Algerian dialect

[3] https://catalog.ldc.upenn.edu/LDC2011T11

mainly borrows words from the French language. These borrowed French words could be classified into two categories [16]. The first one concerns French words that are phonologically altered and the second one concerns words that are pronounced exactly as in French. Table 3 illustrates this categorization with four borrowed French words: the first two are pronounced as if they were Arabic words by using the closest Arabic phonemes, while the last two are pronounced as in French. Consequently, the issue of using French phonemes in an Algerian dialect ASR should be handled.

Table 3. Examples of borrowed French words used in Algerian dialect.

Dialect word	**French word**	**Pronunciation**	**English word**
فَامِيلِيَة	famille	/f a m i l j a/	family
كُوزِينَة	cuisine	/k uː z i n a/	kitchen
قَاطُو	gâteau	/g a t o/	cake
كُونِيكسيُون	connexion	/k ɔ n ɛ k s j ɔ̃/	connection

These previous issues concerned some phonetic specificities of Algerian dialect. Concerning the textual corpus, except social network data, it is difficult to find resources for dialect. The issue is that in social network, there is no standardization in the way of writing a word [2]. In addition, to write the words with foreign phonemes, new symbols have been added to the 28 letters of MSA.

What we propose in the following is to use ALASR to recognize Algerian dialect; to do that we should take into account the characteristics of this dialect.

3.1. Acoustic model

Because in Algerian dialect, French words are used, one should modelize the corresponding phonemes in the acoustic model. Table 4 presents the most frequent French phonemes that are used in Algiers dialect. The initial list of 34 MSA acoustic models should be extended by these French phonemes. An ideal approach would be to train or adapt this new set of models on a transcribed corpus of spoken dialect. Unfortunately, such a resource is not available. Because the task of training the acoustic models, on only the 13 French phonemes presented in Table 4 is difficult, we decided to train all the acoustic models corresponding to the 31 French phonemes. These French acoustic models are then added to the 34 original MSA models.
The French acoustic models are trained on 13 hours of ESTER (Evaluation of Broadcast News enriched transcription systems) corpus [15]. Finally, a DNN-HMM model is built by combining French and Arabic acoustic data.

3.2. Language model

The free way of writing the words in the dialect impacts the estimation of the probabilities of the language model. In fact, unlike MSA where linguistic rules and a typographic writing system have to be respected, there is neither standard, nor rules, to write the dialectal words. For example the MSA phrase قَالت لي (*she told me*) can be written as it is pronounced قَالتلي or it could be written as in MSA. To deal with the issue of dialect writing, the authors of PADIC corpus [17] adopted the following writing rules: if a dialectal word does exist in MSA, it must be written such as in MSA, otherwise the word is written as it is pronounced. They extended also the list of MSA characters by the following ones: پ/p/, ڤ/v/ and ڤ/g/.
We interpolated 3 bigram language models, the two first are trained respectively on the Gigaword and on the speech transcripts of the MSA acoustic training data. The third language model is trained on PADIC, which is, to the best of our knowledge, the largest corpus of Algerian dialect. It includes 6400 parallel dialectal sentences aligned with their corresponding MSA translation. The weights of linear interpolation are optimized in such a way that the likelihood is maximal on a development corpus composed by a mixture of 44K words of MSA and Algerian dialect.

Table 4. The French phonemes used in Algerian dialect with an example of dialectal word for each phoneme.

Phoneme	Dialect word	Pronunciation	English gloss
/ɛ/ /e/	مَاسْتَار	/m a s t ɛ R/	master degree
/g/	بَاقَاج	/b a g a ʒ/	luggage
/p/	پُورْتَابْل	/p ɔ r t a b l/	cell phone
/v/	پُوفْوَار	/p u v w a r/	power
/y/	سِكُورِتِي	/s e k y r i t e/	security
/ɑ̃/	سُورْمُون	/s y R m ɑ̃/	surely
/ø/ /œ/ /ə/	شُومُور	/ʃ ɔ m œ r/	unemployed
/ɛ̃/	لْكُوزَا	/l k u z ɛ̃/	the cousin
/ɔ̃/	كُونْتْر	/k ɔ̃ n t r/	against
/ʒ/	جِيسْت	/ʒ y s t/	right

3.3. Dialect vocabulary

The list of words that will constitute the entries of the vocabulary are extracted from PADIC. But, for speech recognition, the pronunciation of each entry is necessary. There is no resource available containing the phonetic representation of Algerian dialect words. However, a recent work [16] proposed a rule-based grapheme-to-phoneme converter for Algerian dialect. We used this method to produce the phonetic representations of the different entries. Finally, the vocabulary is composed of 4390 unique words with an average of 1.03 pronunciation variants per grapheme entry. This vocabulary dedicated to Algerian dialect has been included into the initial MSA vocabulary of ALASR.
Figure 1 gives details about the distribution of the dialectal words. 42% of the words are exclusively from Algerian dialect, the rest are words from MSA. Among these MSA words, 46% have a different pronunciation, in comparison to the MSA word, 17% have the same pronunciation and 37% do not exist in the initial vocabulary of ALASR system.

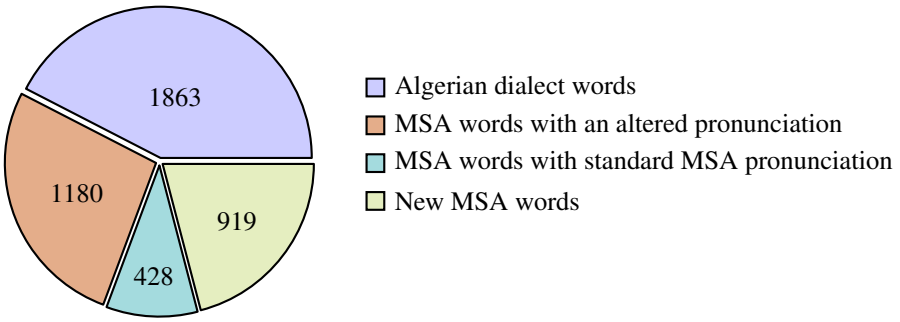

Fig. 1. Distribution of the dialectal lexicon in terms of words.

3.4. Evaluation

Before evaluating ALASR on the dialectal corpus, checking its robustness is desirable, since its acoustic model has been modified to model both MSA and French phones. The objective of this test is to evaluate the impact of the use of a French acoustic model in an ASR system, which has to recognize Arabic and not French. We did an evaluation with this new model on the test corpus on which the original ALASR has been tested. To make the comparison easy, the results of Section 2.5, have been reported in Table 5.

Despite a 1.86% degradation of the word error rate, combining MSA and French phones in the acoustic model leads to acceptable performance on MSA data. Consequently this new version of ALASR can be used on Algerian

Table 5. Performance of ALASR on MSA data before and after combining French and Arabic acoustic data (AM: Acoustic Model, LM: Language Model, Lex: Lexicon).

System	AM	LM	Lex	Dev WER	Test WER
Baseline	MSA	pruned 4-gram	MSA	14.65	15.32
System 1	MSA	3-gram	MSA	13.61	14.52
System 2	MSA	2-gram	MSA	13.54	14.42
Mixture Acoustic Model	MSA+Fr	2-gram	MSA+dialect	14.15	**16.28**

dialect, as the performance on sentences pronounced in MSA is not too much affected, and as the availability of French phones in the acoustic model should help handling words of French origin.

Because no Algerian transcribed speech corpus exists and because it is necessary to evaluate our system on the dialect, the only option is to record a test corpus. That is why 1012 conversational dialect sentences with prepared transcription have been recorded, which corresponds to 1 hour and 10 minutes, by three male Algerian native speakers. This leads to a test corpus composed of 6308 words. It should be noted that the recording was carried out in a quiet recording room using headset microphone and the audio was sampled at 16 kHz.
Tests have been conducted on three variants of ALASR: the baseline with 4-gram language model and without rescoring, system 2 with bigram language model and the one with a mixture of acoustic models. The results are reported in Table 6.

Table 6. Performance of ALASR on Algerian dialect (AM: Acoustic Model, LM: Language Model, Lex: Lexicon).

System	AM	LM	Lex	WER (%)
System 2	MSA	2-gram	MSA	89.04
Baseline	MSA	pruned 4-gram	MSA	87.35
Mixture Acoustic Model	MSA+Fr	2-gram	MSA+dialect	65.45

We can notice that ALASR system collapses completely when tested on the dialect corpus. In fact, it achieves a WER of 89, while on MSA it achieves 14.42. The performance of ALASR with a 4-gram improves slightly the results and achieves a WER of 87.35. By using a mixture of acoustic models (French and MSA), the results are improved substantially (an absolute improvement of 24%). It should be noted that the OOV rate on the test data set is 18%. Even if the results are globally estimated as bad on Algerian dialect in comparison to those achieved for natural language, they could be considered as very encouraging. In fact, in the MGB2017 evaluation campaign, the baseline performance on the Egyptian dialect is of 63.8 [4]. Obviously, we can not compare the two experiments for several reasons (dialects are very different, for the Egyptian, 5 hours of transcribed speech have been used for the training, etc.), but this information is just given, in order to have an idea about the issue of recognizing dialects.

4. Conclusions

In this paper, we presented ALASR, a speech recognition system dedicated to MSA; we evaluated it on MSA. ALASR achieves good results (a WER of 14.02). Then we tested it on Algerian dialect, which is highly influenced by French language. To take into account this specificity, we built a new acoustic model by combining two models: one for MSA and one for French. This combination leads to a WER of 65.45. This result constitutes a real improvement in comparison to the basic system of ALASR, which completely collapses on this data set. It should be mentioned that no transcribed dialectal speech data has been used in the training of the acoustic model and only 4400 sentences have been used with the initial language model. This means that the method we proposed achieves satisfactory results under these extreme conditions of training. In a future work, we will harvest data from social networks to improve the language model and we will also record few hours of data to train the acoustic model.

Acknowledgements

We would like to acknowledge the support of Chist-Era for funding part of this work through the AMIS (Access Multilingual Information opinionS) project.

References

[1] Abdelali, A., Darwish, K., Durrani, N., Mubarak, H., 2016. Farasa: A fast and furious segmenter for arabic, in: HLT-NAACL Demos, Association for Computational Linguistics, San Diego, California.. pp. 11–16.

[2] Abidi, k., Menacer, M.a., Smaili, K., 2017. Calyou: A comparable spoken algerian corpus harvested from youtube, in: Proceedings of Interspeech 2017, p. to appear.

[3] Afify, M., Sarikaya, R., kwang Jeff Kuo, H., Besacier, L., Gao, Y., 2006. On the use of morphological analysis for dialectal arabic speech recognition, in: Proceedings of ICSLP06, pp. 277–280.

[4] Ahmed, A., Stephan, V., Steve, R., 2017. Speech recognition challenge in the wild: Arabic MGB-3.

[5] Al Ameed, H., Al Ketbi, S., Al Kaabi, A., Al Shebli, K., Al Shamsi, N., Al Nuaimi, N., Al Muhairi, S., 2005. Arabic light stemmer: A new enhanced approach, in: The Second International Conference on Innovations in Information Technology (IIT05), pp. 1–9.

[6] Ali, A., Dehak, N., Cardinal, P., Khurana, S., Yella, S., Glass, J., Bell, P., Renals, S., 2016. Automatic dialect detection in arabic broadcast speech. CoRR 08-12-September-2016, 2934–2938. doi:10.21437/Interspeech.2016-1297.

[7] Ali, A., Mubarak, H., Vogel, S., 2014a. Advances in dialectal arabic speech recognition: A study using twitter to improve egyptian ASR, in: International Workshop on Spoken Language Translation (IWSLT 2014).

[8] Ali, A., Zhang, Y., Cardinal, P., Dahak, N., Vogel, S., Glass, J., 2014b. A complete kaldi recipe for building arabic speech recognition systems, in: Spoken Language Technology Workshop (SLT), 2014 IEEE, pp. 525–529. doi:10.1109/SLT.2014.7078629.

[9] Diehl, F., Gales, M.J., Tomalin, M., Woodland, P.C., 2009. Morphological analysis and decomposition for arabic speech-to-text systems, in: Proceedings of Interspeech, pp. 2675–2678.

[10] El-Desoky, A., Gollan, C., Rybach, D., Schlüter, R., Ney, H., 2009. Investigating the use of morphological decomposition and diacritization for improving arabic lvcsr., in: Proceedings of Interspeech, pp. 2679–2682.

[11] El-Desoky, A., Schlüter, R., Ney, H., 2010. A hybrid morphologically decomposed factored language models for arabic lvcsr, in: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics. pp. 701–704.

[12] Elmahdy, M., Hasegawa-Johnson, M., Mustafawi, E., 2012. A baseline speech recognition system for levantine colloquial arabic. Proceedings of ESOLEC .

[13] Elmahdy, M., Hasegawa-Johnson, M., Mustafawi, E., 2014. Development of a tv broadcasts speech recognition system for qatari arabic., in: LREC, pp. 3057–3061.

[14] Gales, M.J., 1998. Maximum likelihood linear transformations for hmm-based speech recognition. Computer speech & language 12, 75–98.

[15] Galliano, S., Gravier, G., Chaubard, L., 2009. The ester 2 evaluation campaign for the rich transcription of french radio broadcasts, in: Proceddings of Interspeech, Brighton (United Kingdom).

[16] Harrat, S., Meftouh, K., Abbas, M., Smaïli, K., 2014. Grapheme to phoneme conversion-an arabic dialect case, in: Spoken Language Technologies for Under-resourced Languages.

[17] Meftouh, K., Harrat, S., Jamoussi, S., Abbas, M., Smaïli, K., 2015. Machine translation experiments on padic: A parallel arabic dialect corpus, in: The 29th Pacific Asia conference on language, information and computation.

[18] Mohri, M., Pereira, F., Riley, M., 2008. Speech recognition with weighted finite-state transducers, in: Springer Handbook of Speech Processing. Springer, pp. 559–584.

[19] Ng, T., Nguyen, K., Zbib, R., Nguyen, L., 2009. Improved morphological decomposition for arabic broadcast news transcription, in: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE. pp. 4309–4312.

[20] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K., 2011. The KALDI speech recognition toolkit, in: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, IEEE Signal Processing Society. IEEE Catalog No.: CFP11SRW-USB.

[21] Soltau, H., Mangu, L., Biadsy, F., 2011. From modern standard arabic to levantine ASR: Leveraging gale for dialects, in: 2011 IEEE Workshop on Automatic Speech Recognition Understanding, pp. 266–271. doi:10.1109/ASRU.2011.6163942.

[22] Stolcke, A., 2000. Entropy-based pruning of backoff language models. arXiv preprint cs/0006025 .

[23] Vergyri, D., Kirchhoff, K., Gadde, V.R.R., Stolcke, A., Zheng, J., 2005. Development of a conversational telephone speech recognizer for levantine arabic, in: Proceedings of Interspeech, Citeseer. pp. 1613–1616.

[24] Veselỳ, K., Ghoshal, A., Burget, L., Povey, D., 2013. Sequence-discriminative training of deep neural networks., in: Proceedings of Interspeech 2013.

[25] Wray, S., Ali, A., 2015. Crowdsource a little to label a lot: Labeling a speech corpus of dialectal Arabic. International Speech and Communication Association. volume 2015-January. pp. 2824–2828.