INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Effects of Dialectal Code-Switching on Speech Modules: A Study using Egyptian Arabic Broadcast Speech

Shammur A Chowdhury[1], *Younes Samih*[1], *Mohamed Eldesouki*[2], *Ahmed Ali*[1]

[1]Qatar Computing Research Institute, HBKU, Doha, Qatar
[2]Concordia University, Montreal, Quebec, Canada
{shchowdhury,amali}@hbku.edu.qa

Abstract

The intra-utterance code-switching (CS) is defined as the alternation between two or more languages within the same utterance. Despite the fact that spoken dialectal code-switching (DCS) is more challenging than CS, it remains largely unexplored. In this study, we describe a method to build the first spoken DCS corpus. The corpus is annotated at the token-level minding both linguistic and acoustic cues for dialectal Arabic. For detailed analysis, we study Arabic automatic speech recognition (ASR), Arabic dialect identification (ADI), and natural language processing (NLP) modules for the DCS corpus. Our results highlight the importance of lexical information for discriminating the DCS labels. We observe that the performance of different models is highly dependent on the degree of code-mixing at the token-level as well as its complexity at the utterance-level.

Index Terms: code-switching, dialect identification, corpus, code mixing index

1. Introduction

With the advent of globalization, code-switching (CS) is increasingly becoming more pervasive in social media and audio/visual media. The phenomena of switching between two or more languages or its varieties is a widespread linguistic phenomenon, analyzed from different interactions, grammatical structures and sociolinguistic perspectives. From the grammatical perspective, CS is categorised into three groups based on its nature of occurrence [1]. The groups are inter-sentential/utterance (between utterance), intra-sentential/utterance (within utterance) and tag switching (a phrase borrowed from another language entity). Among the three, intra-utterance switching is the most problematic for many speech and language processing systems.

Code-switching in spontaneous speech is highly unpredictable and difficult to model. The English-Mandarin language pair has been studied most extensively [2, 3, 4] along with other pairs such as Frisian-Dutch [5], Hindi-English [6, 7] and French-Arabic [8] for CS.

Dialectal code-switching (DCS) remains largely unexplored, especially for spontaneous speech. Even though the recent deep learning models, including automatic speech recognition (ASR) and dialect identification (DI) systems, have achieved groundbreaking results, the presence of dialect switching in the dataset can affect the performance of these speech models. The reason behind such performance degradation can be attributed to the unpredictability of CS points in an utterance, along with the challenge presented by the infusion of the native dialect pronunciation to some non-native words.

From a linguistic and socio-linguistic perspective, the general models for describing CS [9] and the conditions that trigger CS are well researched. Researches have studied CS phenomena in textual datasets for question-answering (QA) [10, 11], language identification [12], name entity recognition (NER) [13] among others. Language modelling is one of the rich researched areas for CS and used most notably to handle CS in machine translation (MT) and automatic speech recogniser (ASR). For processing CS in speech, a variety of approaches have proven to be successful, ranging from applying linguistic knowledge to language-independent methods. Studies such as [14] applied recurrent neural network (RNN) on language models and factored language models to the task of identifying code-switching in speech. To build a better model, the authors in [15] augmented their CS model with syntactic and semantic features. A similar study is done by authors in [2, 16, 17]. For a detailed account of CS in speech, refer to [18].

Because of the scarcity of CS linguistic resources and the prevalence of monolingual data, authors in [19] trained two separate models for the host-guest languages and combined them with a probabilistic model for CS between the two languages. Similarly, authors in [20] proposed to adapt the RNN language model to different CS behaviors and use them to generate artificial code-switching text data. This illustrates the boundaries set to research approaches due to the lack of adequate CS speech datasets, and thus making these challenges hard to explore and address. To the best of our knowledge, there are no studies for dialectal-Arabic code-switching (DACS) in spontaneous speech.

Therefore, we design and develop the first DACS corpus for broadcast speech, annotated at the token-level, considering both the linguistic and the acoustic cues. The dataset contains the following four classes: (*i*) modern standard Arabic (MSA); (*ii*) Egyptian dialect (EGY); (*iii*) MSA in Egyptian accent (MIX) (*iv*) non-Arabic (foreign - FRN) words within a spoken utterance. This dataset is a potential benchmark for DCS in spontaneous speech.

In addition to the dataset design, we analyze the capability of linguistic representation to discriminate between the annotated CS labels. Furthermore, we investigate the effect of DCS on the performance of monolingual state-of-the-art Arabic ASR systems and Arabic dialect identification (ADI) models. We evaluate the performance of the speech models based on the level of code-mixing in utterances. We also analyze how the presence of MIX token impact the models' performance.

In summary, our contributions are: (*i*) define and label CS categories in dialectal speech; (*ii*) release the corpus for the community for further research;[1] (*iii*) analyze the capability of the linguistic representation to discriminate DCS labels; (*iv*) investigate the effect of DCS and intra-word code-mixing on ASR

[1]https://github.com/qcri/Arabic_speech_code_switching

http://dx.doi.org/10.21437/Interspeech.2020-2271

Table 1: *Data description for MGB-3 and DACS dataset. MSA: modern standard Arabic, EGY: Egyptian dialect, MIX: MSA in Egyptian accent, and FRN: foreign words.*

MGB3-EGY		DACS dataset	
Source	MGB-3 dev	# Utterances	1297
Audio Description	16KHz, 16 bit PCM	Annot Unit	word-level
		Vocab	5323
# Files	315	Avg Words/Utt.	11.2
Total Dur	2 hrs	Avg Dur/Utt	5.3 sec
Annot. Labels	EGY	Annot. Labels	MSA/EGY/MIX/FRN
Transcription	from ASR	Transcription	Manual

and ADI models.

2. Arabic Dialectal Code-Switching Dataset

In our study, we use the two-hours Egyptian data from the ADI-5 development dataset in the MGB-3 challenge [21]. The released MGB-3 data includes speech features and textual features extracted from ASR transcription.

Given that the aim is to study the CS between dialectal and standard Arabic and its influence on the performance of other speech models, the models should have access to the verbatim transcription. Consequently, we manually segment the audio into smaller utterances (when there is 500 msec silence or more) and transcribe the speech verbatim by a lay native Egyptian speaker. We opt for the non-linguist transcriber to avoid any bias to standard Arabic.

After transcribing the data, we annotate these segmented utterances for word-level CS information (example of annotation in Figure 1) using the guidelines mentioned in Section 2.1. We then evaluate the annotation (see Section 2.2) to ensure the quality of the dataset and discuss the aggregation technique and distribution in Section 2.3. Details of the dataset is presented in Table 1.

2.1. Annotation Scheme

Based on the observational analysis of the data, we design the annotation guideline for labelling each word in the utterance, while distinguishing between dialectal and standard Arabic. The CS annotation guidelines include the following instruction:

- The corpus was segmented to smaller chunks whenever there is 500 msec silence or more, followed by verbatim transcription.
- Each word should be judged based on the context, and annotators should listen to the audio clip for annotation decision and removing any confusion.
- No modification of the transcription is allowed to be consistent across all annotators.
- Code-switching should not be confused with borrowing [22]. If foreign words are included, the word should be annotated as 'FRN'.
- A 'NULL' tag can be assigned in case the word is unintelligible or cannot be categorised to one of the four labels.

Using the above guidelines, the annotators were asked to classify the words into one of the following four categories: (i) *MSA*: MSA word with MSA pronunciations; (ii) *EGY*: Egyptian word; (iii) *MIX*: MSA word with dialectal pronunciations and (iv) *FRN*: Foreign word, i.e., not Arabic.

Three annotators performed the annotation task. As specified in the guidelines, they annotated each word in the manually transcribed speech utterance.

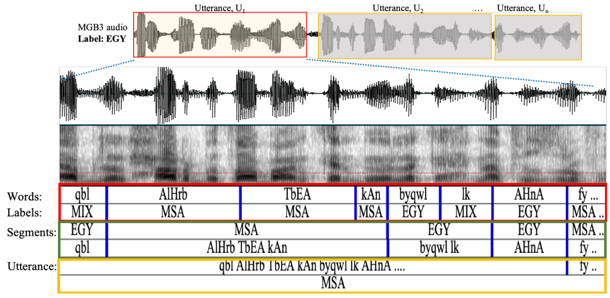


Figure 1: *An example of dialectal code-switching token-level annotation within same utterance. The segments and utterance level labels are inferred from the token-label annotation.*

Table 2: *Inter-Rater Agreement between annotators E, K, and S using Cohen's Kappa Coefficient (CKC) and Fleiss' Kappa Coefficient (FKC).*

	E & K	E & S	K & S	E & K & S
CKC	0.747	0.914	0.802	-
FKC	-	-	-	0.738

2.2. Annotation Evaluation

To assess the reliability of the annotations, we calculate the *inter-rater agreement* using kappa measure, namely *Cohen's Kappa Coefficient* for the agreement between two annotators and Fleiss' Kappa to calculate the agreement among all three annotators. We reported these different agreement values in Table 2. Clearly, it is evident that there is a very high inter-annotation agreement on the word-level labelling, which means there is a pattern to model.

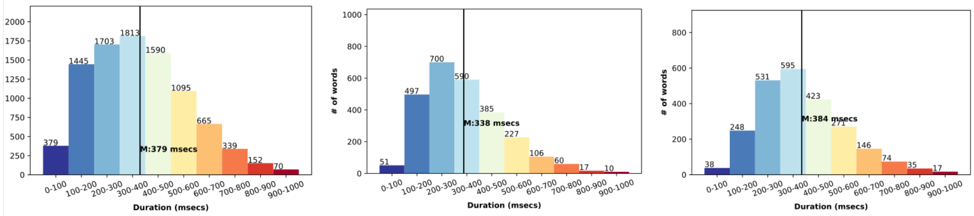
2.3. Data Distribution

As each word has 3 annotation labels, we assign the label agreed by the majority (i.e. at least $\frac{2}{3}$) of the annotators as the final label of the word. Details of annotation distribution per annotator (E, K, S) and majority voting (Maj) are presented in Table 3. Our previous assumption that Egyptian devset in the ADI corpus is based on 100% Egyptian tokens is clearly incorrect. We can see that less than 40% of the words are Egyptian (after assuming MIX is Egyptian) and the rest is MSA. One plausible explanation is that this data was collected in the broadcast domain where it is mainly dominated by formal speech (MSA).

The duration distribution of the tokens in MSA, EGY and MIX are presented in Figure 2. From the figure, we observe that the mean duration of MSA (379 msec) and MIX (384 msec) is significantly more than that of EGY (338 msec). This observation, also including functional words, may indicate the difference in fluency and style of the speakers when speaking in their native dialect than in standard Arabic.

Table 3: *Number of words per annotator (E, K, S) and after word-wise majority voting (Maj) for final labels with its corresponding speech duration. Total given statistic is based on Maj.*

	E	K	S	Maj	H:M:S
EGY	2507	3061	2656	2657	00:14:57
MSA	9538	8484	9113	9307	00:58:49
MIX	2310	2780	2581	2387	00:15:16
FRN	91	114	96	95	00:00:47
Total	Token: 14446 and Duration: 01:29:50				



(i) MSA (ii) EGY (iii) MIX

Figure 2: *Word-Level duration distribution in milliseconds for MSA (Mean:* 379 *ms), EGY (*338 *ms) and MIX (*384 *ms).*

3. Code-Switching and its Consequences

In this section, we provide details of the CS level measures. We then investigate the effect of DCS on processing spontaneous speech, particularly ASR and ADI. Furthermore, we explore the NLP aspects of the DCS, mainly their lexical features: character and word sequences using the verbatim transcription. Finally, we discuss our findings and observations.

3.1. Code-switching Measures

To measure the amount of code-switching in the dataset, we consider both the utterance and corpus level *Code-Mixing Index* (CMI) [23]. We calculate utterance level CMI (C_u) using Equation 1.

$$C_u(x) = \frac{\frac{1}{2} * (N(x) - \max_{L_i \in \mathbf{L}} \{t_{L_i}\}(x)) + \frac{1}{2}P(x)}{N(x)} \quad (1)$$

where N is the number of tokens in utterance x. $L_i \in \mathbf{L}$, the set of all labels in the dataset; $\max\{t_{L_i}\}$ represent the maximum token in the majority label class, with $1 \leq \max\{t_{L_i}\} \leq N$; and P is the number of code alternation points in x; $0 \leq P < N$. We utilize the CMI to evaluate how the increase in the level of code-switching effects the performance of the model.

We report the corpus level CMI by simply averaging the utterance level switching.[2] The details of the level of code-switching in DACS corpus are presented in Table 4.

Table 4: *Details of code-switching level of DACS data using CMI range. The statistics for dataset are after token-label aggregation using majority voting. word/Utt. represents the average word count per utterance, CA is the mean number of code alternation points in utterances, #. presents the number of utterances that belong to that particular CMI range. The dataset does not include examples for 30-45% CMI range.*

Range	word/Utt.	CA	#.
0%	8.98	0.00	121
0-15%	13.34	1.80	107
15-30%	11.44	3.13	260
45-100%	11.18	6.12	809
Corpus CMI	36.5		

Given that the DACS corpus is annotated on the token-level, mainly concerned with the MSA, EGY and MIX labels, this will provide a unique opportunity to investigate the effect of such mixed-code (MIX) token on different speech models. For this study, we calculate the percentage of MIX token, $Pmix = \frac{f(MIX)}{N}$, using frequency, $f(.)$ of MIX token in the segment or utterance of length N words.

[2]However, this does not account for the switches between the utterances

3.2. Speech Models

Data Preparation: We evaluate ASR based on the level of code-switching in each utterance. As for the ADI, we evaluate the model performance using utterance, segment and word-level annotation.

For obtaining utterance-level dialect label, we opt for a majority-based (U_M) approach. We simply consider the total count of tokens labeled as either MSA or EGY[3] and then assign the most frequent label. Using U_M, we obtain 437 EGY and 856 MSA utterances.[4] For segments, we concatenate tokens with the same label (shown in Figure 1).

Automatic Arabic Speech Recognition System: We deploy a grapheme-based acoustic model for the ASR trained using the MGB-2 and MGB-3 data [21, 24]. The recognition experiments are performed using the Kaldi ASR toolkit [25]. We train a conventional context-dependent Gaussian mixture model-hidden Markov model (GMM-HMM) system with 40k Gaussians using 39-dimensional Mel frequency cepstral coefficient (MFCC) features including the deltas and delta-deltas to obtain the alignments. These alignments are used for training a time delay neural network TDNN [26] using sequence discriminative training with the LF-MMI objective [27]. The input to the TDNN is composed of 40-dimensional high-resolution MFCC extracted from frames of 25 msec length and 10 msec shift along with 100-dimensional i-vectors computed from 1500 msec. Five consecutive MFCC vectors and the chunk i-vector are concatenated, forming a 300-dimensional features vector each frame.

For the study, we used two 4-grams language models: (*i*) ASR_MSA, trained using the MGB-2 data [24] with overall word error rate (WER) **45.6%** on the proposed DACS corpus; (*ii*) ASR_EGY, trained using the MGB-3 [21] and Egyptian tweets [28] using transfer learning [29]. The overall WER for ASR_EGY is **42.4%** for the same corpus. We study the performance of both ASR_MSA and ASR_EGY and the results with CS levels were comparable. Therefore, we only report the findings of ASR_MSA system in this paper.

Arabic Dialect Identification Model: We deploy an end-to-end acoustic classifier to distinguish between dialectal-EGY *vs* standard-MSA on the utterance level. We adopt an end-to-end architecture with four temporal convolution neural networks, followed by a global pooling, then passed to two fully-connected layers (1500 and 600 neurons). Rectified Linear Units (ReLUs) are used as activation functions, and the network is trained with stochastic gradient descent (SGD), and 0.001 learning rate. For the training dataset, we use the MGB-2 [24] for MSA instances and EGY validation and test subset from the ADI17 dataset [30, 31]. The overall performance of the trained ADI model using the 315 EGY and 283 MSA dataset[5], is macro $F1 = 63\%$.

[3]MIX mapped to EGY; and other labels are ignored for the U_M.
[4]4 utterances were ignored due to the presence of all FRN words.
[5]From the MGB-3 Dev

3.3. NLP Models

Unlike the previous experiments, using pre-trained speech models, for this task, we train and test the linguistic models using the DACS corpus. We explore different neural architectures including character (char) BiLSTM, word BiLSTM and char-word BiLSTM models to establish strong CS classification results. Our finding is that the char model yielded the best result, as shown in Table 5, similar to previous work in [32]. Therefore, in this study, we only report the architecture for char-BiLSTM model.

For the character based model, we randomly initialize the input embeddings, with d-dimensional ($d = 50$) vectors. We then passed the input embeddings to a bidirectional LSTM (100 units in each direction) layer, followed by a softmax output layer. We train our models using SGD with momentum, optimizing the cross entropy objective function. For the experiments, we use 5-fold cross-validation with identical folds for all experiments.

Table 5: *Reported F-measure with word-char-based features using a simple BiLSTM architecture.*

labels	word	char	char-word
MSA	0.87	**0.91**	0.90
EGY	0.70	**0.82**	0.78

3.4. Model Performance and DCS

Effect on ASR Performance:

For observing how ASR perform with different CS level, we evaluate DACS dataset with different CMI and $Pmix$ range. From Figure 3(a), we observe that the WER increases considerably with increasing CMI value range. A similar pattern is found for sentence error rate.[6] Similarly, from Figure 3(b), we observe that with increase in $Pmix$ in the utterance, WER increases.

Effect on ADI Performance:

The obtained results from the ADI model, for the CMI ranges, are given in Figure 3(a). From the figure, we observe that the increase in CMI-value (i.e. from 15%- onwards), the reported weighted F-measure (W.Avg) significantly decreases. Moreover, we also explore the effect of MIX token on these utterances, segments and word-level performance of the ADI model. In Figure 3(b) (orange and green), we notice that with the increase in $Pmix$, the accuracy of *EGY vs MSA* recognition decreases. This effect is more visible in segment level than the utterance level.

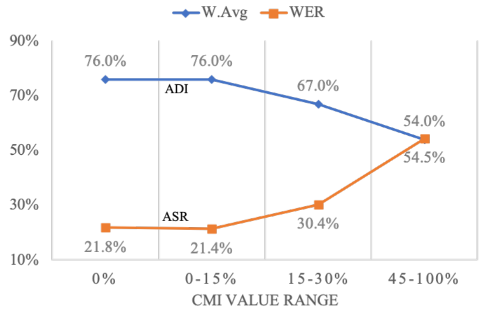
As for word level classification, we notice that even though the MIX token has Egyptian dialectal pronunciation, most of the time it is confused with MSA (76%), thus creating more uncertainty to the ADI model.

Discriminating CS using NLP:

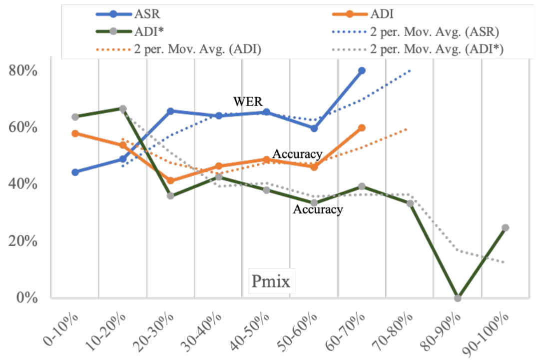
Table 5, reflects that the linguistic representation is highly informative for detecting CS in broadcast dataset. Similar to our studies using $Pmix$, we observe MIX words (mapped as EGY in this paper) are recognized more as MSA. From our manual inspection of the words and its prediction, we found that the most prominent reason for incorrect classification could be attributed to the shared vocabulary between the MSA and MIX, making it difficult for an NLP-based classifier to distinguish between them without the acoustic cues.

Key Observations:

We observe that with increasing code-mixing index score (CMI), the performance of ASR and ADI become considerably poorer. Our findings from $Pmix$ suggest that in addition to CS between words in an utterance, the mixed-code within a word/token can also affect the overall performance of speech processing modules. Our classification using lexical representation indicates its informative capability for predicting with the CS points in an utterance. However, this requires verbatim transcription which is not available in a typical dialectal speech processing scenario.



(a) ASR (WER) and ADI (W.Avg) performance with CMI



Freq.	464	383	247	142	43	13	5	-	-	-	Utt.
	4287	9	78	122	21	317	89	15	2	1221	Seg

(b) Pmix effects on ASR and ADI performance

Figure 3: *Effects of DACS on ASR and ADI model. Figure 3(a) Reported changes in WER for Arabic ASR model and weighted F-measure (W.Avg) for ADI model with different code-mixing index CMI range available in the DACS. Figure 3(b) shows the effect of code-mixing in token-level (MIX) using Pmix on – ASR Utterance level WER, ADI Utterance level and ADI Segment level Accuracy. The number (Freq.) of utterances (Utt.) and segments (Seg.) corresponding to Figure 3(b) is given in the Table with the Figure.*

4. Conclusion

In this study, we build the first spoken dialectal Arabic code-switching (DACS) corpus. The dataset studies code-switching between Egyptian and modern standard Arabic in broadcast domain. We analyze ASR and ADI performance minding the code-mixing index (CMI). We also highlight the importance of NLP information for discriminating dialectal code-switching labels using DACS. For future work, we will increase the dataset with additional dialects and different genres.

[6] For brevity, we are not presenting SER in the paper.

5. References

[1] S. Poplack, "Sometimes I'll Start a Sentence in Spanish y Termino en Espanol: Toward a Typology of Code-Switching," *The Bilingualism Reader*, 2000.

[2] Y. Li and P. Fung, "Improved Mixed Language Speech Recognition Using Asymmetric Acoustic Model and Language Model with Code-Switch Inversion Constraints," in *ICASSP*, 2013.

[3] ——, "Language Modeling for Mixed Language Speech Recognition Using Weighted Phrase Extraction." in *Interspeech*, 2013.

[4] N. T. Vu, D.-C. Lyu, J. Weiner, D. Telaar, T. Schlippe, F. Blaicher, E.-S. Chng, T. Schultz, and H. Li, "A First Speech Recognition System for Mandarin-English Code-Switch Conversational Speech," in *ICASSP*, 2012.

[5] A. Biswas, F. de Wet, E. van der Westhuizen, E. Yilmaz, and T. Niesler, "Multilingual Neural Network Acoustic Modelling for Asr of Under-Resourced English-Isizulu Code-Switched Speech," in *Interspeech*, 2018.

[6] A. Pandey, B. M. L. Srivastava, R. Kumar, B. T. Nellore, K. S. Teja, and S. V. Gangashetty, "Phonetically Balanced Code-Mixed Speech Corpus for Hindi-English Automatic Speech Recognition," in *LREC*, 2018.

[7] K. Taneja, S. Guha, P. Jyothi, and B. Abraham, "Exploiting Monolingual Speech Corpora for Code-Mixed Speech Recognition," *Interspeech*, 2019.

[8] D. Amazouz, M. Adda-Decker, and L. Lamel, "Addressing Code-Switching in French/Algerian Arabic Speech," in *Interspeech*, 2017.

[9] B. E. Bullock and A. J. Toribio, *The Cambridge Handbook of Linguistic Code-Switching*. Cambridge Univ. Press, 2009.

[10] K. Chandu, E. Loginova, V. Gupta, J. v. Genabith, G. Neumann, M. Chinnakotla, E. Nyberg, and A. W. Black, "Code-Mixed Question Answering Challenge: Crowd-Sourcing Data and Techniques," in *Third Workshop on Computational Approaches to Linguistic Code-Switching*, 2019.

[11] K. C. Raghavi, M. K. Chinnakotla, and M. Shrivastava, "" answer ka type kya he?" Learning to Classify Questions in Code-Mixed Language," in *Proceedings of the 24th International Conference on World Wide Web*, 2015.

[12] T. Solorio, E. Blair, S. Maharjan, S. Bethard, M. Diab, M. Ghoneim, A. Hawwari, F. AlGhamdi, J. Hirschberg, A. Chang *et al.*, "Overview for The First Shared Task on Language Identification in Code-Switched Data," in *Proceedings of the First Workshop on Computational Approaches to Code Switching*, 2014.

[13] G. Aguilar, F. AlGhamdi, V. Soto, M. Diab, J. Hirschberg, and T. Solorio, "Named Entity Recognition on Code-Switched Data: Overview of The Calcs 2018 Shared Task," *arXiv preprint arXiv:1906.04138*, 2019.

[14] H. Adel, N. T. Vu, and T. Schultz, "Combination of Recurrent Neural Networks and Factored Language Models for Code-Switching Language Modeling," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2013.

[15] H. Adel, N. T. Vu, K. Kirchhoff, D. Telaar, and T. Schultz, "Syntactic and Semantic Features for Code-Switching Factored Language Models," *Trans. Audio, Speech and Lang. Proc.*, 2015.

[16] Y. Li and P. Fung, "Code-Switch Language Model with Inversion Constraints for Mixed Language Speech Recognition," in *COLING*, 2012.

[17] E. van der Westhuizen and T. Niesler, "Synthesising Isizulu-English Code-Switch Bigrams Using Word Embeddings." 2017.

[18] S. Sitaram, K. R. Chandu, S. K. Rallabandi, and A. W. Black, "A Survey of Code-Switched Speech and Language Processing," *CoRR*, 2019.

[19] S. Garg, T. Parekh, and P. Jyothi, "Dual Language Models for Code Switched Speech Recognition," *arXiv preprint arXiv:1711.01048*, 2017.

[20] N. T. Vu and T. Schultz, "Exploration of The Impact of Maximum Entropy in Recurrent Neural Network Language Models for Code-Switching Speech," in *Proceedings of The First Workshop on Computational Approaches to Code Switching*, Doha, Qatar, 2014, pp. 34–41.

[21] A. Ali, S. Vogel, and S. Renals, "Speech Recognition Challenge in the Wild: Arabic MGB-3," *ASRU*, 2017.

[22] B. L. Hoffer, "Language Borrowing and the Indices of Adaptability and Receptivity," 2005.

[23] B. Gambäck and A. Das, "Comparing the Level of Code-Switching in Corpora," in *LREC*, 2016.

[24] A. Ali, P. Bell, J. Glass, Y. Messaoui, H. Mubarak, S. Renals, and Y. Zhang, "The MGB-2 challenge: Arabic Multi-Dialect Broadcast Media Recognition," in *SLT*, 2016.

[25] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The Kaldi Speech Recognition Toolkit," in *ASRU*, 2011.

[26] V. Peddinti, D. Povey, and S. Khudanpur, "A Time Delay Neural Network Architecture for Efficient Modeling of Long Temporal Contexts," in *INTERSPEECH*, 2015.

[27] D. Povey, V. Peddinti, D. Galvez, P. Ghahrmani, V. Manohar, X. Na, Y. Wang, and S. Khudanpur, "Purely Sequence-Trained Neural Networks for ASR Based on Lattice-Free MMI," in *Interspeech*, 2016.

[28] A. Ali, H. Mubarak, and S. Vogel, "Advances in Dialectal Arabic Speech Recognition: A Study Using Twitter to Improve Egyptian ASR," in *International Workshop on Spoken Language Translation (IWSLT 2014)*, 2014.

[29] S. Khurana, A. Ali, and J. Glass, "DARTS: Dialectal Arabic Transcription System," *arXiv preprint arXiv:1909.12163*, 2019.

[30] S. Shon, A. Ali, Y. Samih, H. Mubarak, and J. Glass, "ADI17: A Fine-Grained Arabic Dialect Identification dataset," in *ICASSP*, 2020.

[31] A. Ali, S. Shon, Y. Samih, H. Mubarak, A. Abdelali, J. Glass, S. Renals, and K. Choukri, "The MGB-5 Challenge: Recognition and Dialect Identification of Dialectal Arabic Speech," in *ASRU*, 2019.

[32] Y. Samih, S. Maharjan, M. Attia, L. Kallmeyer, and T. Solorio, "Multilingual Code-Switching Identification via LSTM Recurrent Neural Networks," in *Proceedings of the Second Workshop on Computational Approaches to Code Switching*, 2016.