Benchmarking Dialectal Arabic-Turkish Machine Translation

Hasan Alkheder hasan.alkhder2@ogr.sakarya.edu.tr
Computer Engineering, Sakarya University, Serdivan, Türkiye

Houda Bouamor hbouamor@cmu.edu
Information Systems, Carnegie Mellon University Qatar, Doha, Qatar

Nizar Habash nizar.habash@nyu.edu
Computer Science, New York University Abu Dhabi, Abu Dhabi, UAE

Ahmet Zengin azengin@sakarya.edu.tr
Computer Engineering, Sakarya University, Serdivan, Türkiye

Abstract

Due to the significant influx of Syrian refugees in Turkey in recent years, the Syrian Arabic dialect has become increasingly prevalent in certain regions of Turkey. Developing a machine translation system between Turkish and Syrian Arabic would be crucial in facilitating communication between the Turkish and Syrian communities in these regions, which can have a positive impact on various domains such as politics, trade, and humanitarian aid. Such a system would also contribute positively to the growing Arab-focused tourism industry in Turkey. In this paper, we present the first research effort exploring translation between Syrian Arabic and Turkish. We use a set of 2,000 parallel sentences from the MADAR corpus containing 25 different city dialects from different cities across the Arab world, in addition to Modern Standard Arabic (MSA), English, and French. Additionally, we explore the translation performance into Turkish from other Arabic dialects and compare the results to the performance achieved when translating from Syrian Arabic. We build our MADAR-Turk data set by manually translating the set of 2,000 sentences from the Damascus dialect of Syria to Turkish with the help of two native Arabic speakers from Syria who are also highly fluent in Turkish. We evaluate the quality of the translations and report the results achieved. We make this first-of-a-kind data set publicly available to support research in machine translation between these important but less studied language pairs.[1]

1 Introduction

The rapid advancements in machine translation technology have significantly helped to break down language barriers and facilitate cross-cultural communication using distant languages, including Turkish and Arabic. Given that Syria and Iraq border Turkey to the south, there is cultural overlap and close ties between those nations and Turkey, making Arabic and Turkish two of the most widely spoken languages in the Middle East. Despite this, there has been no significant research or machine translation effort that specifically addresses the translation of

[1]The MADAR-Turk data set is available from `http://resources.camel-lab.com/`.

Proceedings of Machine Translation Summit XIX, Vol. 1: Research Track, pages 261–271
September 4–8, 2023, Macau SAR, China.

dialectal Arabic and Turkish. The presence of more than 4 million Syrian and Iraqi refugees in Turkey, as well as the massive spread of Turkish drama (dubbed TV series) in the Arab world (Kraidy and Al-Ghazzi, 2013), not to mention the growing tourism industry in Turkey catering to Arab tourists, reinforce the urgent need for developing machine translation capabilities between Turkish and Arabic and its dialects to promote communication and cultural exchange between the Arab countries and Turkey.

Efforts to develop neural machine translation between several language pairs including Turkish (Qumar et al., 2023) and Arabic (Gamal et al., 2022) have yielded promising results, improving translation quality and reducing the need for extensive linguistic knowledge. Building such systems requires a large amount of data, which currently does not exist for the Turkish and Arabic language pair – for Modern Standard Arabic (MSA), and more so for the dialects. Focusing on benchmarking, we present the first research effort exploring translation between Syrian Arabic and Turkish using a set of 2,000 parallel sentences from the MADAR corpus containing 25 different city dialects from various cities across the Arab world, in addition to MSA, English, and French. Additionally, we explore the translation performance into Turkish from other Arabic dialects and compare the results to the performance achieved when translating from Syrian Arabic. We make the data set publicly available.[1]

The paper is structured as follows. Section 2 presents some related work in Turkish and Arabic machine translation, and section 3 discusses the linguistic challenges of Arabic-Turkish translation. Section 4 details the MADAR-Turk data set creation process. Sections 5 and 6 present our benchmarking results and error analysis, respectively. We conclude and describe our future work in Section 7.

2 Related Work

2.1 Arabic-Turkish Resources

Due to the lack of parallel corpora between Arabic and Turkish, MT between this language pair did not receive much attention. A few researchers attempted to develop resources, models, and techniques to translate between these two languages. For instance, Durgar El-Kahlout et al. (2019) introduced an Arabic-to-Turkish statistical machine translation system in the news domains. This work included building parallel Turkish and Arabic corpora collected in different ways: manual translation by professional translators, web-based open-source Arabic-Turkish parallel texts, and using back-translation techniques to translate monolingual Arabic data by using existing machine translation systems. The corpus they created is small and does not include any dialectal Arabic examples.

The OpenSubtitles corpus (Lison and Tiedemann, 2016), a large dataset of TV and movie subtitles covering more than 60 languages, contains Standard Arabic-Turkish parallel texts comprising almost 28 million sentences. Baali et al. (2022) introduced an unsupervised approach to creating a Turkish-Arabic speech corpus from dubbed TV series videos. This corpus was not transcribed and therefore is not available in a text format.

Some research efforts explored Arabic and Turkish for different NLP tasks (Sliwa et al., 2018; Zampieri et al., 2020); however, these efforts employed non-parallel corpora.

A comprehensive survey of the corpora and lexical resources, publicly available for Turkish, is presented in Çöltekin et al. (2023). None of the resources described include dialectal Arabic.

2.2 Dialectal Arabic Parallel Resources

Previous research has focused on creating parallel dialectal data with other languages, but not with Turkish. For instance, MADAR (Bouamor et al., 2018) is the first city-level dialectal dataset including dialects from 25 cities, in addition to MSA, English, and French. MADAR was built on the Basic Traveling Expression Corpus (BTEC) (Takezawa et al., 2007). We draw inspiration from this effort and build on the MADAR corpus to leverage its parallelism benefits in our corpus development. We note that a Turkish version of the BTEC corpus was used for Turkish-English MT (Köprü, 2009; Mermer et al., 2010; Demir et al., 2012); however, to the best of our knowledge, it is not publicly available.

There have been many efforts in Arabic dialect machine translation (Salloum and Habash, 2011; Zbib et al., 2012; Meftouh et al., 2015; Harrat et al., 2017; Baniata et al., 2018; Kchaou et al., 2020; Sghaier and Zrigui, 2020). The work we present in this paper is intended to bridge a crucial gap in the Arabic dialect-Turkish language pairs; we hope this will lay the foundation for further exploration and research in this area.

3 Challenges of Arabic-Turkish Translation

While Ottoman Turkish used to be written in Arabic Script, Modern Turkish uses the Roman script, which adds to the many linguistic differences between Turkish and Arabic and its dialects in terms of morphology, syntax, and lexicon.

3.1 Orthography Differences

Arabic orthography, i.e., the way Arabic language information is encoded using its script, is different from Turkish orthography in the crucial detail of not specifying short vowels and doubling consonants, which are typically written with optional diacritical marks in Arabic. This leads to important ambiguities that pose a significant challenge for Arabic to Turkish MT. For example, the two Arabic words عِقْد *ςiqd*[2] 'necklace' and عَقْد *ςaqd* 'contract' are often written simply as عقد *ςaqd*, but they would be properly translated to Turkish as *kolye* and *sözleşme*, respectively.

3.2 Morphological Differences

Despite centuries of linguistic exchange and geographical proximity, Turkish and Arabic belong to distinct and separate language families. Turkish belongs to the Turkic language family, while Arabic belongs to the Semitic language family. Consequently, there are several morphological differences between the two languages. Most evident is that Arabic is morphologically rich and employs a combination of templatic and affixational morphology (including a number of clitics); while Turkish is heavily agglutinative in nature.

One example of the difference is the absence of the gender feature in Turkish, unlike Arabic's two-gender system. Also Turkish does not have a definite/indefinite distinction. For example, Turkish *büyük sultan* '[a/the] great [male/female] sultan' maps to four Arabic phrases that vary in gender and definiteness: سلطان عظيم *slTAn ςĎym*, السلطان العظيم *AlslTAn AlςĎym*, سلطانة عظيمة *slTAnħ ςĎymħ*, السلطانة العظيمة *AlslTAnħ AlςĎymħ*. We expect this to make mapping from Arabic to Turkish easier than the reverse. The gender neutrality of Turkish even extends to pronouns. For instance, Turkish *o* 'he/she' map to Arabic هو *hw* 'he' and هي *hy* 'she'.

Another important difference is that Arabic utilizes prepositions, but Turkish uses agglutinating postpositions, e.g., the postposition *+a* 'to' *büyük sultana* 'to [the] great sultan'. This compares with the Arabic preposition +ل *l+* in للسلطان العظيم *llslTAn AlςĎym* 'for the great sultan'.

For more information on Arabic and Turkish morphology, see (Habash, 2010) and (Oflazer, 1993).

[2]The Arabic transliteration is in the Habash-Soudi-Buckwalter (HSB) scheme (Habash et al., 2007).

3.3 Syntactic Differences

Syntactically, Turkish is a head-final language that uses a subject-object-verb (SOV) word order; while Arabic is a head-initial language that uses both VSO and SVO orders. For example, the Turkish sentence *çocuk süt içti* '[lit. child milk drank] the child drank milk', is translated as Arabic شرب الطفل الحليب *šrb AlTfl AlHlyb* '[lit. drank the-child the-milk]' or الطفل شرب الحليب *AlTfl šrb AlHlyb* '[lit. the-child drank the-milk]'.

Similarly, Turkish adjectives precede the nouns they modify, while Arabic adjectives follow, as in the Turkish example *büyük sultan* '[the] great sultan' mapping to Arabic السلطان العظيم *AlslTAn AlςĎym* '[lit. the-sultan the-great]', presented above.

Given the complex morphology of both Arabic and Turkish, one can expect many interactions between syntax and morphology in the context of translating between these languages. The examples of Arabic prepositional clitics and Turkish postpositional clitics (shown above) map to separate words when translated: Arabic prepositional clitic +ل *l+* 'for' maps to Turkish standalone postposition *için*, and Turkish postpositional suffix *+a* 'to' maps to the Arabic standalone preposition إلى Ǎlý.

3.4 Lexical Differences and Similarities

Due to the historical and geographical affinity between Arabic and Turkish, there are many words that are shared between the two languages. However, the majority of their lexicons are distinct from each other. Examples of Turkish words of Arabic origin include *kalem* 'pen' from قلم *qalam*, *kahve* 'coffee' from قهوة *qahwaħ*, *merhaba* 'hello' from مرحبا *mrHbA*, and *inşallah* 'God willing' from the phrase إن شاء الله *Ǎn šA' Allh*.

In addition, there are Turkish words that have made their way into standard Arabic such as Turkish *Gümrük* 'customs' and Arabic جمرك *jmrk* and also into dialectal Arabic, particularly Levantine, such as Turkish *Aferin* 'well done' becoming Arabic عفارم *ςfArm*. While the shared lexical items may be useful in translation, in principle, the differences in script, orthography, and morphology can limit their practical value.

3.5 Arabic Dialect Differences

Since we benchmark MT from a number of Arabic dialects, we should note that these varieties differ in many ways at all linguistic levels, including phonology, morphology, syntax, and lexicon (Bouamor et al., 2018; Salameh et al., 2018; Althobaiti, 2020). The differences can even be high within the same country and region. For instance, Salameh et al. (2018) show, as part of their work on dialect identification, that Damascus and Aleppo dialects are different from each other only by 32% and from Beirut dialect by 38%; and that the dissimilarity between the cluster enclosing the Tunisian cities of Tunis and Sfax and the cluster containing the rest of the dialects is more than 50%.

4 MADAR-Turk Data Set Creation

4.1 Data Selection

We used the MADAR Corpus (Bouamor et al., 2018), which was the first set of parallel sentences that include the dialects of 25 Arab cities in addition to English, French, and MSA. Table 1 lists the various cities in the corpus with their countries and regions. MADAR was built on the Basic Traveling Expression Corpus (BTEC) (Takezawa et al., 2007) which comprised about 20,000 English tourism-related sentences. BTEC is conversational in nature, has short sentences, and has translations in several languages, making it an attractive resource for build-

Region	Maghreb				Nile Basin	Levant		Gulf		Yemen
Sub-region	**Morocco**	**Algeria**	**Tunisia**	**Libya**	**Egypt/Sudan**	**South Levant**	**North Levant**	**Iraq**	**Gulf**	**Yemen**
Cities	Rabat Fes	Algiers	Tunis Sfax	Tripoli Benghazi	Cairo Alexandria Aswan Khartoum	Jerusalem Amman Salt	Beirut Damascus Aleppo	Mosul Baghdad Basra	Doha Muscat Riyadh Jeddah	Sana'a

Table 1: The MADAR resources include a variety of region, sub-region, and city dialects.

Turkish	Arabic
Orda, tam turizm ofisinin önünde.	موجود هنيك، قدام مكتب معلومات السياح بالزبط.
Daha önce burda öyle bir adres olduğunu hiç duymadım	ما سمعت بهيك عنوان هون من قبل.
Eczaneyi görene kadar düz git.	إمشي مباشرة لحد ما تشوف صيدلية.
Kahvaltı ne kadar?	بأديش الفطور؟
Sana nasıl yardımcı olabilirim?	كيف فيني ساعدك؟
Soldaki üçüncü aradan geç.	لف عالشمال بالدخلة التالتة.

Table 2: Examples of translation from the Damascus Arabic dialect to the Turkish language.

ing machine translation models. Bouamor et al. (2018) translated large portions of BTEC to five major city dialects representing distinct regions: Beirut (Levant), Doha (Gulf), Cairo (Egypt), Tunis, and Rabat (Maghreb); and they translated a smaller portion (2,000 sentences) to all 25 cities, which plus MSA constitute Corpus-26. In all their translations they started with English or French to avoid the priming effect of Standard Arabic on dialect speakers. In this paper, we work with the same smaller portion and add a Turkish translation to it. This allows us to benchmark translation to Turkish from all Arabic dialects.

4.2 Data Set Construction

Two native Arabic speakers from Syria who are highly fluent in Turkish translated all 2,000 sentences. We provided the translators with a set of guidelines, such as translating each sentence independently without considering the previous context, paying attention to the correctness of the punctuation, and avoiding sentence combinations. After confirming their adherence to these guidelines using an initial pilot set of 50 sentences, the translators proceeded to translate the remainder of the 2,000 sentences from the Damascus dialect into Turkish, from scratch. We specifically chose Damascus because our initial objective was to work on Syrian Arabic to Turkish MT. We expect, and acknowledge, a bias towards Damascus in the effort. Examples of translations from the Damascus dialect into Turkish are shown in Table 2.

4.3 Data Set Statistics

Table 3 presents examples of parallel sentences from the MADAR and MADAR-Turk corpora with their average lengths. We note a stark difference in the number of words per sentence in Turkish (6.9) compared to English (9.9), French (11.5), and most Arabic variants (around 7). This difference is expected due to the agglutinative nature of the Turkish language which results in longer words and fewer overall words per sentence.

Language | Dialect	Example	# words/sentence
Turkish	Eczaneyi görene kadar düz git.	6.9
English	Go straight until you see a drugstore.	9.9
French	Continuez tout droit jusqu'à ce que vous	11.5
MSA	استمر في السير في هذا الطريق حتى تجد صيدلية .	8.0
Aleppo	روح ساوي لبين ما تشوف صيدلية.	6.8
Alexandria	امشى على طول لحد ما تشوف صيدلية.	7.3
Algiers	امشي قبالة حتى تشوف صيدلية.	7.3
Amman	امشي دغري لعند ما تشوف صيدلية.	7.3
Aswan	امشي على طول لغاية متشوف صيدلية.	7.3
Baghdad	اطلع كبل لحد ما تشوف الصيدلية.	6.8
Basra	اطلع كبل لحد ما تشوف صيدلية.	6.6
Beirut	روح دغري لحتى تشوف صيدلية.	6.7
Benghazi	أمشي طول لعند ما تشوف الصيدلية.	7.2
Cairo	أمشي علطول لحد ما تلقي صيدلية.	7.2
Damsacus	إمشي مباشرة لحد ما تشوف صيدلية.	6.8
Doha	امش سيده لين تشوف صيدلية.	6.7
Fes	سير نيشان حتا تلقى الصيدلية.	7.3
Jeddah	امشي سيدا لين ما تلاقي الصيدلية.	6.7
Jerusalem	خليك ماشي دغري لتلاقي صيدلية.	7.0
Khartoum	أمشي دغري لغاية تشوف صيدلية.	7.4
Mosul	امشي كبل الى ان اتشوف صيدليه.	7.1
Muscat	روح سيده حتى تشوف الصيدلية.	7.3
Rabat	سير نيشان حتا تشوف صيدلية.	7.4
Riyadh	امش على طول لين تلقى صيدلية.	7.0
Salt	روح دغري حتى تشوف الصيدلية.	7.1
Sanaa	امشي طوالي لوما تبسر صيدليه.	7.1
Sfax	كمل القدام لين تارى الفارمسي.	6.8
Tripoli	برا طول لين تشوف صيدلية.	7.2
Tunis	امشي طول طول حتى لين تشوف صيدلية.	6.9

Table 3: Examples of parallel sentences from the MADAR and MADAR-Turk corpora with their average lengths.

5 Benchmarking Dialect Arabic to Turkish MT

We translated the various sentences from the MADAR data set into Turkish using *Google Translate*.[3] To evaluate the quality of the automatic translations, we compared them against the reference translations produced by the translators. We measure the translation quality using BLEU (Papineni et al., 2002). We use the SacreBleu implementation (Post, 2018) for evaluating automatic translations against the reference translations (lowercase=True, tokenize='intl'). The results are shown in Table 4. The Table has two parts: (a) organized by the city and (b) organized by region. The results show the following: of all the input languages, MSA has the highest BLEU score, followed by English, then the Riyadh dialect, then French. In contrast, the dialects of Sfax and Tunis (both Tunisian cities) have the lowest scores. Interestingly, English was not the highest, despite its widespread use and the availability of high-quality translation resources. One possible explanation for this result is that we used one reference translation that

[3] `https://translate.google.com/`

(a)

Region	Country	Variant	BLEU
Gulf	Oman	Muscat	23.60
Gulf	Qatar	Doha	20.49
Gulf	Saudi Arabia	Jeddah	20.58
Gulf	Saudi Arabia	Riyadh	**26.92**
Iraq	Iraq	Baghdad	21.46
Iraq	Iraq	Basra	20.25
Iraq	Iraq	Mosul	17.74
Levant	Jordan	Amman	21.84
Levant	Jordan	Salt	22.43
Levant	Lebanon	Beirut	15.81
Levant	Palestine	Jerusalem	22.47
Levant	Syria	Aleppo	21.27
Levant	Syria	Damsacus	26.18
Maghreb	Algeria	Algiers	14.79
Maghreb	Libya	Benghazi	18.54
Maghreb	Libya	Tripoli	16.07
Maghreb	Morocco	Fes	13.64
Maghreb	Morocco	Rabat	9.76
Maghreb	Tunisia	Sfax	8.30
Maghreb	Tunisia	Tunis	8.94
Nile Basin	Egypt	Alexandria	24.13
Nile Basin	Egypt	Aswan	21.95
Nile Basin	Egypt	Cairo	21.99
Nile Basin	Sudan	Khartoum	22.17
Yemen	Yemen	Sanaa	21.09
		MSA	**33.88**
		French	**26.22**
		English	**30.01**

(b)

Region	BLEU
Nile Basin	22.56
Levant	21.67
Yemen	21.09
Gulf	**22.90**
Iraq	19.82
Maghreb	12.86

Table 4: (a) BLEU scores for Google Translate output starting with texts from the various Arab cities in MADAR Corpus, plus Modern Standard Arabic (MSA), English, and French. (b) Average BLEU scores by Arabic dialectal region.

was originally translated from Arabic. The dialect of Damascus was not the best, even though that was the dialect we used when we generated the reference, because the model was developed independently by Google.

We also summarize in Table 4 (b) the differences across the different regions in the Arab world following the regional division that we explained in Table 1. The best performance is in the Gulf, followed by the Nile Basin, followed by the Levant, and the Maghreb appears in the last ranking.

Clearly, a lot more effort has to be done to aid Turkish translation from all these different languages, especially from Arabic.

Language / Dialect	Quality (1- 5)	Transliteration Error	Semantic Error	Morphology Error
English	4.45	4	33	7
French	4.38	2	38	9
MSA	4.26	4	44	8
Damascus	3.80	19	51	22
Tunis	2.74	42	77	30

Table 5: Results of manual evaluation and transliteration, semantic, and morphology errors.

Turkish Reference	Language/ Dialect	Input	Google Translate	Quality (1 - 5)	Transliteration Error	Semantic Error	Morphology Error
Orda, tam turizm ofisinin önünde.	English	It's over there, just in front of the tourist information.	Orada, turist bilgilerinin hemen önünde.	4		X	
	French	C'est là-bas, juste en face du bureau d'information touristique.	Orada, turizm danışma bürosunun hemen önünde.	5			
	MSA	هناك ، أمام بيانات السائح تماما.	Orada, turizm verilerinin hemen önünde.	4		X	
	Damascus	موجود هنيك، قدام مكتب معلومات السياح بالزبط.	Orada, turizm danışma bürosunun hemen önünde.	4		X	
	Tunis	اهوكا غادي، بالضبط قدام البيرو متاع الارشادات السياحية.	Ahuka Gadi, Peru'nun hemen önünde, turist rehberlerinin keyfini çıkarın.	1	X	X	
Bu biraz soğuk. Isıtabilir misin?	English	This is rather cold. Will you heat it up?	Bu oldukça soğuk. ısıtacak mısın?	5			
	French	C'est plutôt froid. Pouvez-vous le réchauffer?	Oldukça soğuk. Bunu ısıtabilir misin?	5			
	MSA	هذا بارد إلى حد ما . هل يمكن أن تقوم بتسخينه ؟	Bu biraz havalı. Isıtabilir misin?	3		X	
	Damascus	هاد بارد شوي. ممكن تسخنه؟	Bu soğuk. Isıtabilir misin?	5			
	Tunis	هذا بارد شوية. بش تسخنو؟	Bu biraz soğuk. Isınıyor musun?	3			X
Teşekkür ederim. Bu yeterli.	English	Thank you, that's enough.	Teşekkürler, bu kadar yeter.	5			
	French	Merci. Ce sera tout.	TEŞEKKÜRLER. Bu kadar.	5			
	MSA	شكراً ، هذا يكفي.	Teşekkürler, bu kadar yeter.	5			
	Damascus	شكراً. هيك بكفي.	teşekkür ederim. Bu yeterli.	5			
	Tunis	شكرا، يزي.	Teşekkürler, Yeezy.	3	X		

Table 6: Examples from the manual error analysis.

6 Error Analysis

In addition to the quantitative evaluation using BLEU, we conducted an error analysis on translations from the several languages we studied, specifically English, French, and MSA because these are standard languages, as well as the dialect of Damascus and Tunis (which was among the worst-performing in the evaluation).

We chose the same 100 sentences for these languages and evaluated their Turkish automatic translation outputs in two different ways. Firstly, we asked human evaluators to rate the translation quality on a scale of 1 to 5, where 5 represents a perfectly acceptable translation in Turkish that accurately covers the meaning and fluency of Turkish, and 1 represents a translation that is lacking in either accuracy or fluency in a way that makes it hard to read and has errors.

Additionally, we identify three types of errors: transliteration errors, semantic errors, and morphology errors. **Transliteration** errors refer to cases where the system failed to translate a word and produced a transliteration instead, e.g., Tunisian Arabic يزي *yzy* ‘enough’ is transliterated as *Yeezy* instead of Turkish *yeterli*. **Semantic** errors refer to cases where a word is translated with a different meaning than intended. For instance, the Damascus Arabic word نص (with ambiguous diacritization as *nuS∼* 'half' or *naS∼* 'text') is incorrectly translated in the context of the phrase نص قنينة *nS qnynħ* 'half bottle' as *şişe metni* 'bottle text' as opposed to the correct translation *yarım şişe* 'half bottle'. And **morphology** errors refer to cases where a word is translated with errors in morphological features. For example, the Tunisian Arabic verb نحب *nHb* 'I want' (Turkish reference *istiyorum*) is mistranslated as *seviyoruz* 'we love' (i.e. plural instead of singular morphology). This is most likely a result of confusion with the MSA reading of the Arabic word which also means 'we love'.

The summary of our results is given in Table 5. We provide examples in Table 6. English has the highest quality; which is expected given that it is a language with a wealth of resources and training data. Furthermore, we observe that, despite being the best-automated automated assessment using BLEU, MSA came in third place in terms of translation quality behind English and French. Lastly, the Tunisian dialect had the lowest quality and had the greatest errors compared to the other languages evaluated.

7 Conclusion and Future Work

We introduced MADAR-Turk, a set of 2,000 sentences from the MADAR corpus, translated from the Damascus dialect into Turkish. To the best of our knowledge, this is a first-of-a-kind human reference set for Dialectal Arabic-Turkish. Our study provides the first-ever benchmarking results on translation performance from Arabic dialects to Turkish. By producing this data set and making it publicly available, we hope to support ongoing efforts to improve translation and language access for individuals who speak Arabic dialects in the Turkish context.

In the future, we plan to continue expanding the human reference set to improve machine translation in the context of this resource-scarce language pair. We also plan to use this data set as part of developing improved methods for machine translation for low-resource language pairs.

Acknowledgments

We would like to express our sincere gratitude to Sara Almarmour and Ahad Hızıroğlu for their exceptional effort in translating the dataset used in this study. We are truly grateful for their assistance. We also thank the reviewers for their valuable feedback and comments.

References

Althobaiti, M. J. (2020). Automatic Arabic dialect identification systems for written texts:A survey. *CoRR*, abs/2009.12622.

Baali, M., El-Hajj, W., and Ali, A. (2022). Creating speech-to-speech corpus from dubbed series. ArXiv preprint arXiv:2203.03601.

Baniata, L. H., Park, S., Park, S.-B., et al. (2018). A neural machine translation model for arabic dialects that utilizes multitask learning (mtl). *Computational intelligence and neuroscience*, 2018.

Bouamor, H., Habash, N., Salameh, M., Zaghouani, W., Rambow, O., Abdulrahim, D., Obeid, O., Khalifa, S., Eryani, F., Erdmann, A., and Oflazer, K. (2018). The MADAR Arabic dialect corpus and lexicon. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Çöltekin, Ç., Doğruöz, A. S., and Çetinoğlu, Ö. (2023). Resources for turkish natural language processing: A critical survey. *Language Resources and Evaluation*, 57(1):449–488.

Demir, S., El-Kahlout, İ. D., Unal, E., and Kaya, H. (2012). Turkish paraphrase corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 4087–4091, Istanbul, Turkey. European Language Resources Association (ELRA).

Durgar El-Kahlout, İ., Bektaş, E., Erdem, N. Ş., and Kaya, H. (2019). Translating between morphologically rich languages: An Arabic-to-Turkish machine translation system. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pages 158–166, Florence, Italy. Association for Computational Linguistics.

Gamal, D., Alfonse, M., Jiménez-Zafra, S. M., and Aref, M. (2022). Survey of arabic machine translation, methodologies, progress, and challenges. In *2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pages 378–383. IEEE.

Habash, N., Soudi, A., and Buckwalter, T. (2007). On Arabic Transliteration. In van den Bosch, A. and Soudi, A., editors, *Arabic Computational Morphology: Knowledge-based and Empirical Methods*, pages 15–22. Springer, Netherlands.

Habash, N. Y. (2010). *Introduction to Arabic natural language processing*, volume 3. Morgan & Claypool Publishers.

Harrat, S., Meftouh, K., and Smaïli, K. (2017). Machine translation for arabic dialects (survey). *Information Processing & Management*.

Kchaou, S., Boujelbane, R., and Hadrich-Belguith, L. (2020). Parallel resources for Tunisian Arabic dialect translation. In *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, pages 200–206, Barcelona, Spain (Online). Association for Computational Linguistics.

Köprü, S. (2009). AppTek Turkish-English machine translation system description for IWSLT 2009. In *Proceedings of the 6th International Workshop on Spoken Language Translation: Evaluation Campaign*, pages 19–23, Tokyo, Japan.

Kraidy, M. M. and Al-Ghazzi, O. (2013). Neo-ottoman cool: Turkish popular culture in the arab public sphere. *Popular Communication*, 11(1):17–29.

Lison, P. and Tiedemann, J. (2016). Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 923–929. European Language Resources Association.

Meftouh, K., Harrat, S., Jamoussi, S., Abbas, M., and Smaili, K. (2015). Machine translation experiments on padic: A parallel arabic dialect corpus. In *Proceedings of the 29th Pacific Asia conference on language, information and computation*, pages 26–34.

Mermer, C., Kaya, H., and Doğan, M. U. (2010). The TÜBİTAK-UEKAE statistical machine translation system for IWSLT 2010. In *Proceedings of the 7th International Workshop on Spoken Language Translation: Evaluation Campaign*, pages 183–188, Paris, France.

Oflazer, K. (1993). Two-level description of Turkish morphology. In *Sixth Conference of the European Chapter of the Association for Computational Linguistics*, Utrecht, The Netherlands. Association for Computational Linguistics.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the Conference of the Association for Computational Linguistics (ACL)*, pages 311–318, Philadelphia, Pennsylvania, USA.

Post, M. (2018). A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Qumar, S. M. U., Azim, M., and Quadri, S. (2023). Neural machine translation: A survey of methods used for low resource languages. In *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 1640–1647. IEEE.

Salameh, M., Bouamor, H., and Habash, N. (2018). Fine-grained Arabic dialect identification.

Salloum, W. and Habash, N. (2011). Dialectal to Standard Arabic Paraphrasing to Improve Arabic-English Statistical Machine Translation. In *Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*, pages 10–21, Edinburgh, Scotland.

Sghaier, M. A. and Zrigui, M. (2020). Rule-based machine translation from tunisian dialect to modern standard arabic. *Procedia Computer Science*, 176:310–319. Knowledge-Based and Intelligent Information and Engineering Systems: Proceedings of the 24th International Conference KES2020.

Sliwa, A., Ma, Y., Liu, R., Borad, N., Ziyaei, S., Ghobadi, M., Sabbah, F., and Aker, A. (2018). Multi-lingual argumentative corpora in english, turkish, greek, albanian, croatian, serbian, macedonian, bulgarian, romanian and arabic. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Takezawa, T., Kikui, G., Mizushima, M., and Sumita, E. (2007). Multilingual Spoken Language Corpus Development for Communication Research. *Computational Linguistics and Chinese Language Processing*, 12(3):303–324.

Zampieri, M., Nakov, P., Rosenthal, S., Atanasova, P., Karadzhov, G., Mubarak, H., Derczynski, L., Pitenis, Z., and Çağrı Çöltekin (2020). Semeval-2020 task 12: Multilingual offensive language identification in social media (offenseval 2020).

Zbib, R., Malchiodi, E., Devlin, J., Stallard, D., Matsoukas, S., Schwartz, R., Makhoul, J., Zaidan, O. F., and Callison-Burch, C. (2012). Machine translation of Arabic dialects. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 49–59, Montréal, Canada. Association for Computational Linguistics.