International Journal of Arabic-English Studies (IJAES) *Vol.24, No.1, 2024*
DOI: https://doi.org/10.33806/ijaes.v24i1.560

The Negative Transfer Effect on the Neural Machine Translation of Egyptian Arabic Adjuncts into English: The Case of Google Translate

Rania Al-Sabbagh
University of Sharjah, UAE

Received: 29.1.2023 Accepted: 16.7.2023 Published: 2.1.2024

Abstract: Parallel corpora for low-resource Arabic dialects and English are limited and small-scale, and most neural machine translation models, including Google Translate, rely mainly on parallel corpora of standard Arabic and English to train for dialectal Arabic translation. A model well trained to translate to and from standard Arabic is believed to efficiently translate dialectal Arabic, given their similarities. This study demonstrates the impact of not using large-scale, dialect-specific parallel corpora by quantitatively and qualitatively analyzing the performance of Google Translate in translating Egyptian Arabic adjuncts. Compared to human reference translation, Google Translate achieved a low BLEU score of 14.69. Qualitative analysis showed that reliance on standard Arabic parallel corpora caused a negative transfer problem manifested in the literal translation of idiomatic adjuncts, the misinterpretation of dialectal adjuncts as main clause constituents, the translation of dialectal adjuncts after orthographically similar standard Arabic words, and the use of standard Arabic common lexical meanings to translate dialect-specific adjuncts. This study's findings will be relevant for researchers interested in dialectal Arabic neural machine translation and has implications for investment in the development of large-scale, dialect-specific corpora to better process the peculiarities of Arabic dialects and reduce the effect of negative transfer from standard Arabic.

Keywords: Arabic dialects, Egyptian Arabic, Google Translate, negative transfer, low-resource dialects, neural machine translation, parallel corpora

1. Introduction

Numerous studies have addressed machine translation between standard Arabic and other languages, especially English. Several rule-based, phrase-based, and neural machine translation models have been trained and tested for a variety of genres. The output of public machine translation models such as Google Translate™ has been extensively analyzed by researchers such as Alkhawaja, Ibrahim, Ghanim and Awwad (2019) and Diab (2021), who argued that the output of Google Translate is of acceptable accuracy and fluency, and that Google Translate can be integrated into the workflow of translation service providers as an economical option. In such a workflow, clients' texts are first translated by Google Translate and then post-edited by professional translators (i.e., translators with formal training in translation, extensive experience with paid translation services, complete mastery of the source and target languages, and full knowledge of the cultural differences between the source and target audiences). The extensive research on standard Arabic machine translation is supported by a wealth of parallel corpora (i.e., repositories of translation equivalents) that have been used to train and test models. However, the lack of similar corpora for dialectal Arabic hinders research on machine translation for Arabic dialects.

Arabic dialects are the native languages used by millions of Arabs in daily conversations, media productions, and social media posts. Standard Arabic is the official written and formal spoken standard used in Arab countries in settings including academia, literature, national and international media, official documents, science and technology, and diplomacy. To facilitate cross-cultural communication, machine translation models should be able to translate not only standard Arabic, but also Arabic dialects. However, machine translation research into dialectal Arabic is still a young field of study, hampered by the scarcity of parallel corpora. Creating such corpora is a time-consuming, labor-intensive, and expensive process—and thus, researchers usually opt for other approaches. The first approach synthesizes dialectal Arabic parallel corpora using seed parallel corpora of standard Arabic and English as well as seed bilingual lexicons of standard and dialectal Arabic (Hassan, Elaraby and Tawfik 2017). The second approach uses standard Arabic as a pivot language so that Arabic dialects are first translated into standard Arabic and then the standard Arabic output is translated into English (Baniata, Park and Park 2018; Farhan, Talafha, Abuammar, Jaikat, Al-Ayyoub, Tarakj and Toma 2020; Baniata, Ampomah and Park 2021; Slim, Melouah, Faghihi and Sahib 2022; Kchaou, Boujelbane and Hadrich 2023). The third approach uses code-mixed parallel corpora that consist of standard and dialectal Arabic sentences (Nagoudi, Elmadany and Abdul-Mageed 2021) so that the standard Arabic portion can leverage the overall performance of the model. The inclusion of standard Arabic in the training of neural machine translation models for Arabic dialects assumes that the similarities between standard and dialectal Arabic would enable machine translation models to efficiently translate the latter. However, the highest bilingual evaluation understudy (BLEU) score for machine translation of Arabic dialects is 27.91 points, achieved by a neural machine translation model developed by Baniata et al. (2018) that translates Levantine Arabic into English using synthetic data. Conversely, the BLEU scores for standard Arabic range from 32.07 (Nagoudi, Elmadany and Abdul-Mageed 2022) to 49.7 (Almahairi, Cho, Habash and Courville 2016). Clearly, despite the similarities between standard and dialectal Arabic, the differences cause machine translation models for dialectal Arabic to lag behind.

This study demonstrates the negative effects of not using large-scale, dialect-specific parallel corpora to train neural machine translation models for Arabic dialects through a quantitative and qualitative analysis of the results of Google Translate for the translation of Egyptian Arabic adjuncts in a corpus of 280 song lyrics. The study's findings will be relevant for researchers interested in dialectal Arabic neural machine translation and has implications for investment in the development of large-scale, dialect-specific corpora to better process the peculiarities of Arabic dialects and reduce the effect of negative transfer from standard Arabic.

Although there are several publicly available machine translation models that support Arabic, such as Microsoft® Translator and Yandex®, Google Translate was chosen because it is the most widely used translation service, with an estimated more than one billion daily users (Pitman 2021). In 2016, Google Translate switched from phrase-based statistical machine translation to neural machine translation, which has been praised for producing fluent output, effectively using contextual information, and requiring minimal human intervention. In this study, the results refer to the neural version of Google Translate.

Egyptian Arabic was chosen because it is the native dialect of the most populated Arab country, the Arab Republic of Egypt (Egypt), whose population is estimated at more than 112 million (United Nations Population Funds 2023). Another reason for choosing Egyptian Arabic is its unique linguistic structures and word meanings, which are influenced by the Coptic language spoken in Egypt before the Arab conquest and a variety of languages spoken by large foreign communities that lived in Egypt for decades, including Italian, Greek, French, and Turkish communities. The uniqueness of Egyptian Arabic is evident in several linguistic constituents, from which I chose to work on adjuncts. Halliday and Matthiessen (2014: 154) defined adjuncts as dispensable constituents that, when removed, have no effect on the structure or meaning of the rest of the sentence. Adjuncts are classified as circumstantial, interpersonal, and textual, depending on the metafunction they serve in the clause. Adjuncts differ greatly between standard and Egyptian Arabic at the lexical and syntactic levels. For example, the textual conjunctive adjunct و (wa) means “and” in standard Arabic, but in Egyptian Arabic it means “and,” “but,” or “when.” In Egyptian Arabic, interpersonal vocative adjuncts can be realized as nominal groups or relative clauses, whereas in standard Arabic they are realized only as nominal groups. These and other differences, discussed in Section 4, cause Google Translate to fail in translating many Egyptian Arabic adjuncts because it has not been sufficiently trained on Egyptian Arabic-specific corpora.

Few parallel corpora of Egyptian Arabic and English exist, as discussed in Section 2.3. The textual genres represented in these corpora include discussion forums (Chen, Tracey, Walker and Strassel 2019), telephone conversations (Kumar, Cao, Cotterell, Callison-Burch, Povey and Khudanpur 2014), text messages (Li, Grimes and Strassel 2019), travel-related questions and answers (Zbib, Malchiodi, Devlin, Stallard, Matsoukas, Schwartz, Makhoul, Zaidan and Callison-Burch 2012; Bouamor, Habash, Salameh, Zaghouani, Rambow, Abdulrahim, Obeid, Khalifa, Eryani, Erdmann and Oflazer 2018), and informal interviews (Hamed, Habash, Abdennadhar and Vu 2022). However, some of these corpora are code-mixed—for example, Egyptian Arabic is mixed with standard Arabic (as in the forum discussions) or with English (as in the informal interviews). In others, the translations were done by amateur translators (i.e., translators with a good command of the source and target languages, but little to no formal translation training or experience with paid translation services) as in telephone conversations and travel-related questions and answers. Additionally, most of these corpora are available by paid subscription.

This study considered the textual genre of song lyrics and the development of a new parallel corpus of Egyptian Arabic song lyrics translated by professional translators. Song lyrics were selected for several reasons. First, the selected songs are the lyrics of a famous Egyptian singer named Amr Diab, whose songs are exclusively in Egyptian Arabic and are not mixed with either standard Arabic or English. Second, song lyrics belong to a creative genre, in which songwriters intentionally break with language conventions and coin new words and clause structures for aesthetic purposes. Therefore, song lyrics are a challenge for Google Translate, which is known for its poor performance in noisy texts—that is, texts with abundant vocabulary and grammatical variations. Finally, song lyrics have not been featured in previous corpora; therefore, this study contributes a new resource for the small repository of parallel Egyptian Arabic corpora with 280 song lyrics translated by professional translators with a total of 3,066 translated segments.

2. Literature review

2.1 Neural machine translation for Arabic dialects

Since 2016, when Google Translate moved from statistical to neural machine translation, neural machine translation has become the de facto paradigm. In this section, I focus on neural machine translation for Arabic dialects including Egyptian Arabic. However, interested readers can refer to Harrat, Meftouh and Samili (2019) and Zakraoui, Saleh, Al-Maadeed and Alja'am (2021) for an overview of rule-based, lexicon-based, phrase-based, and statistical machine translation models for Arabic varieties.

As mentioned in Section 1, the small number of parallel corpora of Arabic dialects is a major obstacle to the development of efficient neural machine translation models, especially because these models require large corpora with millions of translated segments. To overcome this obstacle, researchers usually opt for other approaches. In the first approach, Hassan et al. (2017) synthesized a parallel corpus of Levantine Arabic by using a parallel seed corpus of standard Arabic and English as well as a bilingual seed lexicon of Levantine and standard Arabic. First, each English word was matched with its standard Arabic translation. Second, a k-nearest neighbors algorithm was used to find the synonyms and near-synonyms for each English word and match them with their standard Arabic equivalents. Third, they searched the bilingual lexicon for the Levantine Arabic equivalents of the standard Arabic translations. Finally, they replaced the standard Arabic words in the translated segments with their Levantine Arabic equivalents. When the synthesized corpus was used to fine-tune a large, publicly available neural machine translation model, similar to Google Translate, translation performance improved compared to not using the synthesized corpus, from 25.03 BLEU points to 27.91 points.

The second approach to overcoming the lack of parallel dialectal Arabic corpora is to use standard Arabic as a pivot language. In this approach, Arabic dialects are first translated into standard Arabic and then into English. Baniata et al. (2018) followed this approach and translated Levantine and Maghrebi Arabic into standard Arabic using a multitask neural algorithm. The BLEU score for translating

Levantine Arabic into standard Arabic was 41; however, the score for translating Maghrebi Arabic into standard Arabic was 30. When the standard Arabic output was translated into English, BLEU score further decreased to 27, which is almost the same as the score of Hassan et al. (2017). Baniata et al. (2018) have argued that errors carried over from the translation of dialectal Arabic into standard Arabic to the translation of standard Arabic into English, and that there are many differences between Arabic dialects and standard Arabic that have led to a deterioration in performance. For example, many words in Levantine and Maghrebi Arabic do not overlap with standard Arabic because they are loanwords. Levantine and Maghrebi Arabic are more agglutinative than standard Arabic, so a word like ماجلهاش (majalhāsh) is equivalent to the three-word clause لم يأت لها (lam yaʔti: laha:, he did not come to her). There are morphemes that are used in Arabic dialects but not in standard Arabic, such as the progressive prefixes ba- and ka- in Levantine and Maghrebi Arabic, respectively. There are several standard Arabic sounds that are either replaced by lighter sounds or reduced, such as ثمن (thaman, price), which is typically pronounced and written as تمن (taman) in Arabic dialects, with /th/ replaced by /t/, and انتم (ʔantum, you.PL) in standard Arabic, which is typically pronounced and written as نتو (ntu:) in Maghrebi Arabic.

In one variation on the second approach, Farhan et al. (2020) used a zero-shot approach to build a machine translation model that can translate from Levantine to standard Arabic. A zero-shot approach means that the source Arabic dialect—in this case Levantine Arabic—is not represented in the training corpus. For the training, parallel corpora of Egyptian and Saudi Arabic were created by extracting 309,000 standard Arabic sentences from the OpenSubtitles corpus (Lison and Tiedemann 2016) and translating them into Egyptian and Saudi Arabic. No information was provided about the translation process except that it was performed by a data provider and reviewed by a language expert. The BLEU score for the translation of Levantine–standard Arabic was 31.06 when the model was trained on the Saudi–standard Arabic corpus compared to 24.42 when the model was trained on the Egyptian–standard Arabic corpus. Such a difference might suggest that there are more lexical and syntactic similarities between Levantine and Saudi Arabic than between Levantine and Egyptian Arabic. However, the results of translating the standard Arabic output into English were not reported.

In a second variation on the second approach, Slim et al. (2022) have argued that there are similarities not only between standard and dialectal Arabic but also within Arabic dialects. Therefore, they used the MADAR (Multi Arabic Dialect Applications and Resources) corpus (Bouamor et al. 2018) and the Parallel Arabic Dialect (PADIC) corpus (Meftouh, Harrat, Jamoussi, Abbas and Smaili 2015) to train and test a transfer-based neural machine translation model that translates Algerian Arabic into standard Arabic. MADAR contains parallel sentences from 25 Arabic dialects including Algerian Arabic, and PADIC represents two Algerian dialects in 6,400 sentences along with three other Arabic dialects. The model achieves a BLEU score of 35.87; however, the results for translating the standard Arabic output into English were not reported.

In two other variations on the second approach, recent studies by Baniata et al. (2021) and Kchaou et al. (2023) have obtained much higher BLEU scores when translating Arabic dialects into standard Arabic. However, neither study reports results when translating the standard Arabic output into English. Baniata et al. (2021) applied a new neural word piece model to address a major problem in translating Arabic dialects into standard Arabic, namely the words that do not overlap between dialectal and standard Arabic. The model decomposes each dialectal Arabic word into its morphemes and uses this information to find the closest possible standard Arabic equivalent. In a test of translating Levantine, Egyptian, and Gulf Arabic into standard Arabic, the BLEU scores were 63.71, 48.19, and 47.26, respectively. Kchaou et al. (2023) also achieved a high BLEU score of 60 in translating Tunisian into standard Arabic using a transformer model and a large synthesized parallel corpus of Tunisian and standard Arabic.

A third approach to overcoming the lack of dialectal Arabic parallel corpora is the use of code-mixed corpora. Nagoudi et al. (2021) created a sequence-to-sequence transformer model to translate a code-mixed corpus of standard and Egyptian Arabic. The corpus consisted of (a) 61 million standard Arabic–English sentence pairs extracted from the Open Source Parallel Corpus (OPUS) (Tiedemann 2012), (b) 56,000 Egyptian Arabic translated segments, (c) 160,000 Levantine Arabic translated segments, and (d) 40,700 translated segments from Qatari Arabic. The corpus is code mixed in the sense that it contains some clauses in standard Arabic and others in Egyptian Arabic, but not in the sense that standard and Egyptian Arabic occur together in the same clause. The reported BLEU score is 25.7, which is lower than the scores obtained by Hassan et al. (2017) and Baniata et al. (2018). The authors explained that their test corpus contained more standard Arabic clauses than Egyptian Arabic clauses—this implies that the scores might have been lower if the test corpus contained only Egyptian Arabic clauses to translate.

2.2 Negative transfer in neural machine translation

In neural machine translation, transfer learning involves training a model on a language pair with many parallel corpora (the parent model) and then transferring the parameters learned in the parent model to a language pair with few resources (the child model) for initialization and training. This approach usually leads to a negative transfer problem (Dabre, Chu and Kunchukuttan 2020; Wang, Lipton and Tsvetkov 2020; Richardson and Wiles 2022; Zhang, Deng, Zhang and Wu 2023), which is similar to the problem people face when learning a second language: the first language (in this case, the resource-rich language) interferes with learning or processing of the second language (the resource-poor language). Studies have addressed the problem of negative transfer in neural machine translation from a technical point of view to propose model architectures that could overcome the problem. The proposed architectures include the use of hierarchical knowledge distillation (Saleh, Buntine and Haffari 2021), hyper-adapters (Baziotis, Artetxe, Cross and Bhosale 2022), and adversarial neural networks (Wang, Dai, Póczos and Carbonell 2019; Sun, Wang, Pasquine and Hameed 2021). However, none of these studies adopt a linguistic perspective to describe how negative transfer manifests in translation output. Moreover, most studies attempt to reduce negative transfer between different languages rather than between variants of the same language. Even the studies that work with corpora containing Arabic work on standard not dialectal Arabic and report their overall BLEU scores without specifying which language pair the improvements apply to.

2.3 Parallel corpora of Egyptian Arabic

Some parallel corpora of Arabic dialects and English exist, but not all of them contain Egyptian Arabic. In this section, I reviewed the corpora that contain Egyptian Arabic, and then briefly reviewed parallel corpora for other Arabic dialects. One of the oldest parallel corpora is the Arabic Parallel Text by Zbib et al. (2012), which contains Levantine Arabic–English and Egyptian Arabic–English parallel texts consisting of 1.1 million and 380,000 words, respectively. To create the corpus, monolingual Arabic texts were first collected from weblogs and online user groups. Second, amateur translators were hired from the crowdsourcing platform Amazon Mechanical Turk®. When the corpus was used to train and test a statistical translation system, it scored 20.66 and 19.29 BLEU points for Egyptian Arabic–English and Levantine Arabic–English translation, respectively.

Similar to Zbib et al. (2012), Kumar et al. (2014) hired amateur translators to translate the CALLHOME Egyptian Arabic corpus (Gadalla, Kilany, Arram, Yacoub, El-Habashi, Shalaby, Karins, Rowson, MacIntyre, Kingsbury, Graff and McLemore 1997), its supplement (Linguistic Data Consortium 2002a), and the 1997 Arabic HUB5 evaluation corpus (Linguistic Data Consortium 2002b), with a total of 160 unscripted telephone conversations containing 226,962 words. The corpora were collected as part of the Broad Operational Language Translation (BOLT) program of the U.S. Defense Advanced Research Projects Agency. To ensure translation quality, each utterance was translated four times by four different translators.

Two other corpora created for the BOLT program by the Linguistic Data Consortium are the BOLT Arabic Discussion Forum Parallel corpus (Chen et al. 2019) and the BOLT Egyptian Arabic–English SMS/Chat Corpus (Li et al. 2019). The discussion forum corpus consists of 1,169,599 words, most of which are in Egyptian Arabic. In examining the corpus, I found a significant number of standard Arabic sentences. The SMS/Chat corpus consists of 349,414 Egyptian Arabic words. Both corpora were translated by professional translators. Corpora from the Linguistic Data Consortium are available by paid subscription.

Other corpora containing Egyptian Arabic include the Multidialectal Parallel Corpus of Arabic (MPCA) (Bouamor, Habash and Oflazer 2014), the Multi Arabic Dialect Applications and Resources (MADAR) corpus (Bouamor et al. 2018), and ArzEn-ST corpus (Hamed et al. 2022). MPCA contains 2,000 sentences derived from Zbib et al.'s (2012) corpus. However, the sentences were translated by professional rather than amateur translators. MADAR consists of 2,000 sentences extracted from the Basic Travel Expression Corpus (Takezawa, Kikui, Mizushima and Sumita 2007) and professionally translated into 25 Arabic dialects, including

Egyptian Arabic. The ArzEn-ST corpus is a three-way speech translation corpus of code-switched Egyptian Arabic and English collected through informal interviews with students at an upscale private university in Egypt. The corpus was primarily intended to reflect different types of code-switching: (a) inter-sentential, in which some whole sentences were in Egyptian Arabic and others in English; (b) intra-sentential, in which some words in the same sentence were in Egyptian Arabic and others in English; and (c) morphological, in which the Egyptian Arabic morphemes were agglutinated into English words, as in “bi-t-save” (it saves). All sentences with all code-switching types were translated into English by professional translators, yielding a total of 6,216 sentences. Some of these corpora are available upon request, but at time of the writing, I have been unable to obtain any.

3. Method

3.1 Data

3.1.1 Data collection and cleaning

For the corpus of this study, song lyrics were compiled from https://aghanilyrics.com/, a website where fans volunteer to write the lyrics of songs they like. A search of the website for songs by Amr Diab yielded 280 songs from 1983 to 2020. As mentioned in Section 1, Amr Diab was selected because his songs are exclusively in Egyptian Arabic. The compiled song lyrics were not preprocessed except for removing duplicate lines. Table 1 shows the corpus statistics.

Table 1. Corpus statistics

No. of song lyrics	280
No. of lines (translated segments)	3,066
No. of word tokens	17,848
No. of word types	5,487

3.1.2 Data translation

Four translators and one reviewer translated the compiled corpus. All were native speakers of Egyptian Arabic and well acquainted with the songs. The translators each had four years and the reviewer six years of experience with Egyptian Arabic–English subtitling for Netflix®. The corpus was divided equally among the four translators, and then the reviewer checked the translations for accuracy and consistency. The translation guidelines were as follows:

- Translators must provide natural translations that reflect the intended meaning.
- Translators must avoid literal translations of idioms, figures of speech, and culture-specific references.
- Translators must use mainstream English without slang.
- Translators must maintain American English spelling.
- Translators should ignore so-called singability in translations, i.e., translations need not follow rhyme or rhythm patterns.
- Translators should account for blurbs such as repetition in their translations.

- Translators should keep the punctuation marks of the source text in the same relative position.
- Google Translate should not be used under any circumstances, even if the output has been post-edited. Translations must be done from scratch by professional translators.
- Translators should listen to the song to decipher ambiguous words.

Since Google Translate is updated regularly, it is important to mention when the translations were made, as the results may differ even in the short term. For this study, all translations were created between May 2 and May 8, 2023, using the Google Translate automated programming interface (API).

3.2 Analysis methods and models

3.2.1 Systemic functional grammar

As mentioned in Section 1, Egyptian Arabic differs from standard Arabic in many aspects, and adjuncts are a linguistic constituent with great discrepancy. In this study, I define and classify adjuncts following Halliday and Matthiessen’s (2014) systemic functional grammar. In systemic functional grammar, adjuncts are defined as dispensable clause constituents that, when removed, do not affect the structure or the meaning of the remaining clause. Adjuncts are divided into circumstantial, interpersonal, and textual adjuncts. Table 2 shows the definition of each category along with examples from the corpus.

3.2.2 Bilingual evaluation underscore (BLEU) for quantitative analysis

BLEU is the most popular automatic metric for evaluating machine translation models. It indicates how similar the machine translation is to the human reference translation, with scores closer to 100 indicating more similar translations. In this study, I used Tilde’s (2023) online implementation of BLEU, which is user-friendly and requires no programming experience. I applied BLEU at two levels: for the whole text and for adjunct translation. The former considers the translation of all lines, the latter focuses on how well adjuncts were translated. The use of BLEU makes the results of this study more easily comparable to the studies in Section 2.3. It also provides an overview of how well Google Translate performs. However, it does not provide deep insight into the nature of the errors and their causes. For this reason, I combined the results from BLEU with a manual qualitative analysis.

3.2.3 Qualitative analysis

All adjuncts whose Google Translate translation receives a score of less than 100 BLEU (i.e., whose translations do not perfectly match the human reference) were manually coded to decide whether the error could be the result of negative transfer. The following are considered translation errors that are the result of negative transfer:

- A word that occurs in both standard and Egyptian Arabic, has an additional Egyptian Arabic-specific adjunctive meaning, but Google Translate provides the standard Arabic meaning, not the Egyptian Arabic one:

  - for example, خلاص (khala:ṣ) is a common abstract noun in standard Arabic meaning salvation, and although it can also be used in Egyptian Arabic as salvation, its far more common use in Egyptian Arabic is a textual, continuative adjunct best translated as "that's it," "finally," or "already";
  - if Google Translate translates khalāṣ as salvation, in an Egyptian Arabic clause, this is an error resulting from the negative transfer.
- A word that exists only in Egyptian Arabic, so Google Translate transliterates it rather than translates it:
  - for example, أتاري (ʔata:ri:) is a textual conjunctive adjunct used to mark a conclusion, as in 'atārih kaddāb kibīr (She turned out to be a great liar) and there is no similar word in standard Arabic;
  - if Google Translate cannot translate it correctly and instead renders a transliteration, this can be considered an example of negative transfer.
- An idiomatic Egyptian Arabic-specific adjunct that does not exist in standard Arabic and is translated literally by Google Translate:
  - for example, من بعيد لبعيد (min biᶜi:d libᶜi:d) in Egyptian Arabic is a circumstantial adjunct of manner meaning "secretly";
  - if Google Translate provides a literal translation as "from far to far" because there is no similar idiomatic expression in standard Arabic, this is a translation error caused by negative transfer.
- An Egyptian Arabic adjunct translated after an orthographically similar but lexically irrelevant standard Arabic word:
  - for example, the circumstantial adjunct of comparison زي الجناين (zayy iljana:yin, like gardens) was translated by Google Translate as "like fetuses";
  - it seems that Google Translate missed الجنانين (ʔiljana:yyin, gardens) for the orthographically similar, but lexically different, standard Arabic word أجنة (ʔajinnah, fetuses).
- An adjunct translated as a main constituent of a clause (i.e., subject, predicate, or complement):
  - for example, in the clause أنا عمري من الليلة ابتدى (ʔana: ᶜumri: min illi:lah ʔibtada:), the subject of the verb ابتدى (ʔibtada:, begun) is عمري (ᶜumri:, my life), not أنا (ʔana:, I);
  - the pronoun here is just a textual continuative adjunct to catch the listener's attention;
  - ideally, 'anā should not be translated, and this is exactly what human translators have chosen to do;
  - in standard Arabic, however, 'anā is almost always the subject of the sentence, rarely a discourse filler;
  - thus, when Google Translate translates the clause as "I'm old enough tonight," this is a negative transfer error, since the standard Arabic rules were followed here.

Table 2. Adjuncts, definitions, and examples

Adjunct	Definition	Examples
Circumstantial	They provide background information about the process expressed in the clause, such as time, place, distance, duration, frequency, means, quality, comparison, degree, cause, contingency, accompaniment, role, and point of view.	Time دلوقتي باحلم عندها dilwaʔti: ba-hlam ^{c}anda-ha: now PROG.dream.1SG at-her Now, my dream has come true. Place كان قلبي هنا عندك مرهون ka:n ʔalb-i: hina ^{c}anda-k marhu:n was.3SG.M heart-my here at-you captivated My heart was caught here in your web. Frequency ياما اتمنيت تبقى معايا ya:ma: ʔtmani:t tibʔa ma^{c}a:-yā always wished.1SG become.2SG.M with-me I have always wished to be with you. Comparison زي الزمان مالكيش أمان zayy iz-zama:n ma-li-ki:-sh ʔama:n like the-time not-for-you.2SG.F-not trust Like time, you are fickle and can't be trusted. Cause أنا ياما اتحملت عشانك ʔana: ya:ma: ʔithamilt ^{c}asha:na-k I always endured.1SG for-you I have always endured for your sake. Contingency ولولا حبك هعيش لمين؟ wi-lu:la: huba-k ha-^{c}i:sh li-mi:n? and-without love-your will-live.1SG for-whom? Without your love, I have no other reason to live. Point of view بالنسبالي أعز الناس bi-l-nisba-l-i: ʔa^{c}zz ʔin-na:s as-the-case-for-me dearest the-people For me, she is my sweetheart.

Interpersonal	They express (a) the speaker’s degree of certainty about the statement being spoken; (b) the speaker’s wishful thinking, (c) the speaker’s judgment of or attitude toward the content of the message; (d) the status of the statement, whether affirmative or negative; and (e) the nature of the relationship between the speaker and the listener through vocatives.	Degree of certainty هي أكيد مش واحدة عادية hiyya ʔaki:d mish waḫdah ᶜa:diyyah she definitely not girl ordinary She is definitely not an ordinary girl. يمكن أكون غلطان ومش داري yimkin ʔaku:n ghalṯa:n wi-mush da:ri: maybe be.1SG wrong but-not knowing Maybe I’m wrong, but I don’t know. والله حبك كل عمري wa-allahi ḫuba-k kul ʔumr-i: and-God love-you all life-my I swear to God, your love is my life. Wishful thinking ياريت أنا أشوفك قصادي ya:ri:t ʔana: ʔashu:fa-k ʔuṣa:d-i: wishing I see.1SG-you before-me I wish I could see you before me. Speaker’s judgment/attitude قالك ندم ʔal-la-k nidim said.3SG.M-to-you regretted.3SG.M Sarcastically, she regretted it. Negation معرفش مين كان السبب ma-ᶜraf-sh mi:n ka:n ʔis-sabab not-know.1SG-not who was.3SG.M the-reason I don’t know who is to blame for it. ما نسيت ويارتني نسيت ma nisi:t wi-ya:ri:t-ni: nisi:t not forgot.1SG and-wish-I forgot.1SG I haven’t forgotten, and I wish I had. Vocatives حبيبي جاية أجمل سنين habi:b-i: jayyah ʔajmal sini:n love-my coming.SG.F best years The best years of our lives are still ahead of us, baby.

		قول يااللي كلامك باين في عينيك ʔu:l ya-lli: kala:ma-k bayin fi-ʻinī say.2SG.M oh-who words-your showing in-eye Tell me! Your eyes say it all! (Changed into an independent clause by human translators)
Textual	They are divided into conjunctive and continuative adjuncts. The former maintain cohesion across clauses, and are similar to conjunctions in traditional grammar. The latter—also known as discourse fillers— signal movement in the discourse, whether by the same speaker or a new one.	Conjunctive tu:l ma-nta ẖabi:b-i: ʔana: mush length that-you love-my I not ẖa-nsa:-k will-forget.1SG -you مهما هقول برضو شوية mahma ha-ʔu:l bardu shuwayya regardless will-say.1SG still little No matter how hard I try, I don't have the words to describe her. كل ده علشان بتهرب م الحقيقة kul dah ᶜalasha:n bi-tihrab all this because prog.escaping.2SG.M m il-ẖaʔi:ʔah from the-truth And all because you're running away from the truth. Continuative (in many cases deleted by human translators) أنا قلبي حبك ʔana: ʔalb-i: ẖaba-k I heart-my loved.3SG.M-you My heart has fallen in love with you. ما تسيبك منهم ma:-tisi:bak min-hum it.is.that-leave.2SG.M of-them Let go of them! ده انت في عينيا كل اللي ليا da ʔinta fi-ᶜinayy-a: kul it.is.that you.2SG.M in-eyes-my all ʔilli li-yā that for-me In my eyes, you are my everything.

4. Results

A total of 5,398 adjuncts were extracted, distributed as shown in Table 3. The BLEU score for the whole corpus was 13.55 and for the adjunct translations 14.69. Table 3 also shows the BLEU score for each type of adjunct. These scores are significantly lower than those obtained by Nagoudi et al. (2021) for Egyptian Arabic (see Subsection 2.1). However, a major difference between these results and those of Nagoudi et al. (2021) is that the corpus of this study consists exclusively of Egyptian Arabic, whereas the test corpus used by Nagoudi et al. (2021) contained a mixture of standard and Egyptian Arabic clauses. BLEU scores below 30 typically mean that the output is incomprehensible to users of the target language.

Table 3. Adjunct statistics and BLEU scores

Adjunct	Total Count	BLEU Score
Circumstantial	1,957	17.45
Interpersonal	1,303	9.4
Textual	2,138	15.86

Based on manual coding, at least 20% of adjunct translation errors can be attributed to negative transfer from standard Arabic. Cases of lexical, orthographic, and syntactic transfer were found, as shown in the following subsections.

4.1 Lexical negative transfer

Lexical negative transfer manifests itself in words used in standard and Egyptian Arabic that have similar meanings, but also have Egyptian Arabic-specific, adjunctive meanings. Google Translate has consistently ignored the Egyptian Arabic-specific, adjunctive meanings, as shown by the examples in Table 4 and in خلاص (khalāṣ), mentioned earlier in Subsection 3.2.3.

Table 4. Examples of Google Translate ignoring Egyptian Arabic-specific, adjunct-related meanings

Adjunct	Adjunct Type	Standard/ Egyptian Arabic Meaning	Egyptian Arabic-Specific, Adjunctive Meaning	Human Translation (HT) vs. Google Translate (GT)
عمر ᶜumr	interpersonal (negative polarity)	life/age	never	عمر الخوف مايرد قدر ᶜumr il-khu:f ma:-yirud ʔadar never the-fear not-stops.3SG.M fate But fear never stops fate. *(HT)*

				The age of fear is as much as possible. *(GT)*
موت mu:t	circumstantial (degree)	death	very much	والله لاحبك موت wa-llah la-ẖibi-k mu:t and-God to-love.1SG -you deeply I swear I'll love you like crazy. *(HT)* I swear to God, your love is death. *(GT)*
أوقات ʔawʔa:t	circumstantial (frequency)	times	sometimes	البعد أوقات بيهون ʔil-bu^cd ʔawʔa:t biyhawin the-distance sometimes comforts.3SG.M Being far away sometimes makes it easier *(HT)* Distance is easy times. *(GT)*
حبة ẖabah	circumstantial (manner)	a pill	gradually	واخدني حبة حبة wi-ʔakhad-ni: ẖabah ẖabah and-took.3SG.M-me little little She has gradually captured my heart *(HT)* And he took me a pill. *(GT)*
يا ... يا ya:...ya:	textual (conjunctive)	oh (a vocative particle)	either...or	يا أنا يا لأ ya: ʔana: ya: laʔ either me or not It's either me or no one else *(HT)* Oh me oh no *(GT)*
ما ma:	textual (continuative)	not	after all	ما أنا خدت عليك ma: na khadt ᶜali:-k after.all I accustomed.1SG on-you After all, I've grown accustomed to you. *(HT)* What I took from you *(GT)*

Another way in which lexical negative transfer is manifested is through Egyptian Arabic-specific adjuncts, which do not exist in standard Arabic. A common example of such adjuncts is yāmā, a circumstantial adjunct of frequency, duration, or degree, depending on the context. In Example 1, ya:ma: stands for a long duration; therefore, it is best translated as "for a long time." In Example 2, it stands for intensity and is therefore best translated as "very much." Google Translate often did not recognize ya:ma: as a circumstantial adjunct and transliterated it as if it were a proper noun, sometimes with a capital Y.

1 عليكي قلبي ياما دور
^{c}ali:-ki: Ɂalb-i: ya:ma: dawwar
on.you.3SG.F heart-my long searched.3SG.M
My s has been searching for you for a long time. *Human Translation*
You have my heart, Yama turn. *Google Translate*

2 اشتقنا ليكم ياما
ɁishtaɁna: li:-kum ya:ma:
missed.1PL for-you.PL lots
I've missed you so much. *Human Translation*
We got you Yama. *Google Translate*

Lexical negative transfer also frequently shows up in the literal translation of idiomatic Egyptian Arabic adjuncts. These adjuncts are idiomatic in the sense that they are not compositional: their meanings are not the sum of the meanings of the individual words. These adjuncts are specific to Egyptian Arabic and are not used as adjuncts in standard Arabic. For example, the verbal group قالك (Ɂallak) in Table 2 literally means "he said to you" and it always means that in standard Arabic. However, in Egyptian Arabic it is used as an interpersonal adjunct of disbelief, which means that the speaker does not believe the following statement. Therefore, human translators have translated it as "sarcastically."

A similar idiomatic Egyptian Arabic-specific adjunct is the nominal group أصلها (Ɂaṣlaha:). Literally, it means "its/her origin," and this is the only meaning in standard Arabic. However, in Egyptian Arabic it is often used as a continuative adjunct at the beginning of a sentence to catch the listener's attention, see Example 3. It can be translated as "it is that," but human translators have almost always ignored it because it is a discourse filler that does not contribute to the meaning of the clause. Google Translate has not made a similar translation decision and has always translated it as "its/her origin."

3 أصلها بتفرق في حياتك
Ɂaṣlah bi-tifriɁ fi ẖaya:t
a: : a-k
it.is.th PROG.changes.3S i life-
at G.F n your
Only a woman can change life . *Human Translation*
Its origin makes a difference in your life. *Google Translate*

Idiomatic expressions were very common in interpersonal vocative adjuncts; this is probably why interpersonal adjuncts received the lowest BLEU score, as shown in Table 3. The corpus contains romantic songs in which the life companion is often called out. However, ẖabi:bi: (my lover/love/baby) is not the only way to call out a lover. Other common vocatives are listed in Table 5 along with their literal translations from Google Translate and their correct human translations. In addition to the examples in Table 5, there is an interesting example of idiomatic vocatives used in Egyptian Arabic, but not in standard Arabic, and that is يا سلام (ya: sala:m). It is not exactly calling someone or something, but indicates surprise, astonishment, or disgust, depending on the tone and context. Human translators have rendered it as “Oh my!” and “O God!”; however, Google Translate has consistently translated it as “peace be upon him.”

Table 5**.** Examples of idiomatic Egyptian Arabic interpersonal vocative adjuncts translated literally by Google Translate

Adjunct	Google Translate Literal Translation	Human Translation
يا روحي (ya: ru:ẖi:)	o my soul	darling, my love, my beloved, baby, dear, honey
يا روح الروح (ya: ra:ẖ ʔirru:ẖ)	o soul of the soul	
يا حياتي (ya: ẖaya:ti:)	o my life	
يا عمري (ya: ᶜumri:)	o my age	
يا نور العين (ya: nu:r ʔilᶜi:n)	o light of the eye	o apple of my eyes

Idiomatic Egyptian Arabic-specific expressions are also found in circumstantial adjuncts. One example is من بعيد لبعيد (min bi‘īd lib‘īd, secretly), literally translated as “from far to far,” as mentioned earlier in Subsection 3.2.3. Further examples are given in Table 6.

Table 6. Examples of idiomatic Egyptian Arabic circumstantial adjuncts translated literally by Google Translate

Adjunct	Google Translate Literal Translation	Human Translation	The whole line with Human Translation (HT) and Google Translation (GT)
في يوم وليلة (fi: yu:m wililah)	in a day and a night	overnight	وهعدي في يوم وليلة wi-ha-ᶜadi: fi: yu:m wi-lilah and-will-cross.1SG in day and-night It’ll be done overnight. *(HT)* And I promise in a day and a night. *(GT)*
من غير إزاي وليه (min ghīr ’izzay wilīh)	without how and why	completely	سلمت القلب ليك من غير إزاي وليه salimt ʔil-ʔalb li:-k min ghi:r ʔizzay wi-li:h

			gave.1SG the-heart to-you from without how and-why I completely surrendered my hear to you. *(HT)* I delivered the heart to you without how and why. *(GT)*
بالعربي (bilʿarabī)	in Arabic	frankly and simply put	بالعربي محدش يستاهل bi-l-ᶜarabi: ma-ḥad-sh yista:hil in-the-Arabic no-one-no deserves.3SG.M Frankly and simply put, no one is worth it. *(HT)* In Arabic, no one deserves it. *(GT)*

4.2 Orthographic negative transfer

In many cases, Egyptian Arabic adjuncts have been translated after orthographically similar standard Arabic words, even though these standard Arabic words are completely irrelevant. One example was جناين (jana:yyin, gardens) mentioned in Subsection 3.2.2. It was translated as if it were أجنة (ʔajinnah, fetuses). Another frequently occurring adjunct that was affected by orthographic negative transfer is the circumstantial adjunct of accompaniment وياك (wayya:k, with you). Almost every time wayyāk was translated as “and you” as if it were وإياك (waʔiyya:k, and you). The phrase waʔiyya:k is specific to standard Arabic and is never used in Egyptian Arabic. In Example 4, Google Translate translates the circumstantial adjunct قوام (ʔawa:m, fast) as “strength” as if it were قوة (quwwah). Similarly, the circumstantial adjunct قصادي (ʔuṣa:di:, in front of me) in Example 5 has been translated as “my intention,” which is written as قصدي (qaṣdi:).

4 كبر قوام واتعلم
kibir ʔawa:m wi-tᶜalim
grew.3SG.M fast and-learned. 3SG.M
She grew up fast and learned her lesson. *Human Translation*
Grow strength and learn. *Google Translate*

5 ياريت انا أشوفك قصادي
ya:ri:t ʔana: ʔashu:fa-k ʔuṣa:d-i:
hoping I see.1SG-you front-me
I wish I could see you before me. *Human Translate*
I hope I see you my intention. *Google Translate*

4.3 Syntactic negative transfer

Syntactic negative transfer showed up in textual continuative adjuncts, especially دا (dah, this) and أنا (ʔana:, I). In both standard and Egyptian Arabic, dah is a singular masculine demonstrative pronoun; and ’anā is a singular first-person subject pronoun. However, in Egyptian Arabic, both can also be used as continuative adjuncts at the beginning of a clause to catch the listener’s attention (see Examples 6–9). In general, the professional translators in this study ignored the translation of these adjuncts because they have no syntactic function and do not contribute to the overall meaning of the clause. However, Google Translate consistently mistranslated them as the subject of the clause.

6 يا روحي أنا زادت في قلبي الغيرة

ya:	ru:ḥ-i:	ʔana:	za:dit	fi:	ʔalb-i:	ʔil-ghi:rah
oh	soul-my	I	increased.3SG.F	in	heart-my	the-jealousy

My love, my heart becomes more and more jealous. *Human Translation*
Oh my soul, I increased jealousy in my heart. *Google Translate*

7 أنا مكنش الحب عمره في نيتي

ʔana:	ma-kan-sh	ʔil-ḥub	ᶜumru-h	fi:	niyit-i:
I	not-3SG.M- not	the-love	never-it	in	intention-my

I never thought that I would fall in love. *Human Translation*
I did not love his life in my intention. *Google Translate*

8 ده أنا ياما زمان دوقت الأحزان

da	ʔana:	ya:ma:	zama:n	duʔt	ʔil-ʔaḥza:n
this	I	lots	past	tasted.1 SG	the-sorrows

For a long time, I have tasted sorrow. *Human Translation*
This is me, the time of sorrows. *Google Translate*

9 ده مهما انجرحنا بتدبل جروحنا

da	mahma:	ʔinjaraḥ-na:	bi-tidbal	jru:ḥ-na:
it.is.that	regardless	hurt-PASS-PAST-us	PROG-heal.3SG.F	wounds-our

Whatever wounds are created, they always heal. *Human Translation*
This is no matter how bad we hurt our wounds. *Google Translate*

5. Discussion

The low BLEU score of Google Translate and the fact that more than 20% of translation errors are due to lexical, orthographic, and syntactic negative transfer from standard Arabic show the importance of including more dialect-specific parallel corpora in the training of neural machine translation models to better account for the specifics of each dialect. Google Translate is a massive model that has been trained over years on billions of translated segments—yet it performed poorly for Egyptian Arabic because it did not have enough training segments.

There are several ways to create parallel corpora. The first, used in this study, is to collect monolingual data and have them translated by experts. This is the most reliable method, but it is also costly and time-consuming. A second option is crowdsourcing (i.e., hiring amateur translators on the Internet), as done by Zbib et al. (2012) and Kumar et al. (2014), mentioned in Subsection 2.3. This method is much cheaper and less time-consuming. However, the quality of the translation is always questionable because the hired amateur translators do not necessarily have translation skills and there is always a risk of fraudsters, e.g., people using Google Translate for translation and submitting it as their own translation. Zbib et al. (2012) and Kumar et al. (2014) used techniques to detect fraudsters, such as measuring the similarity between Google Translate and the submitted translations and excluding those with high similarity, and manually checking samples of the submitted translations. However, such techniques do not necessarily guarantee high quality. A third method for creating parallel corpora is to create synthetic data, similar to the work of Baniata et al. (2018) and Kchaou et al. (2023) mentioned in Subsection 2.1. Synthesizing data is rapid and inexpensive because it relies on automated modules and requires only seed corpora and bilingual lexicons. However, the data is still not as authentic as natural data.

Regardless of how parallel corpora should be created, building such corpora for Arabic dialects and testing the extent to which they can improve the performance of current neural machine translation models are rich areas of research that remain to be explored, especially for Egyptian Arabic. This study provides empirical evidence for further research in these areas by showing how the lack of dialect-specific parallel corpora can lead to poor machine translation performance.

6. Conclusion

This study has shown that the limited availability of dialect-specific parallel corpora hinders the ability of neural machine translation models such as Google Translate to accurately process Arabic dialects such as Egyptian Arabic, even when these models have been extensively trained with vast amounts of parallel corpora of standard Arabic and English. When translating a corpus of Egyptian Arabic songs into English, Google Translate achieved a BLEU score of only 14.69, indicating that the output is barely comprehensible to the target audience. Moreover, manual review of the translations of Egyptian Arabic adjuncts revealed that at least 20% of translation errors were due to lexical, orthographic, and syntactic negative transfer from standard Arabic. Idiomatic adjuncts were translated literally. Dialect-specific adjunctive meanings were ignored. Some adjuncts were transliterated instead of translated. Some other adjuncts were translated after orthographically similar but lexically irrelevant standard Arabic words. Continuative adjuncts were mistranslated as main clause constituents, especially subjects. Therefore, this study argues that researchers interested in neural machine translation should invest in creating more dialect-specific corpora and that adding these corpora to the standard Arabic-English parallel corpora used for training could help the models process dialect-specific features. To this end, the study contributes a new parallel corpus of Egyptian Arabic song lyrics that have been translated into English by professional translators and is available upon request. However, there are two limitations to the current study. First, negative transfer was described to manifest only in adjuncts; however, negative transfer can manifest in many other linguistic constituents. Second, it has been reported that at least 20% of errors in the output of Google Translate are due to negative transfer from standard Arabic; such a percentage may vary depending on the genre and dialect studied. Therefore, further studies are needed to investigate the different manifestations of negative transfer in a range of linguistic constituents, text genres, and Arabic dialects.

Rania Al-Sabbagh, PhD
Department of Foreign Languages
University of Sharjah
ORCID number: 0000-0002-1208-5115
Email: rmalsabbagh@sharjah.ac.ae

References

Alkhawaja, Linda, Hanan Ibrahim, Fida' Ghnaim and Sirine Awwad. (2019). 'Neural machine translation: Fine-grained evaluation of Google Translate output for English-to-Arabic translation'. *International Journal of English Linguistics*, 10 (4): 43–60.

Almahairi, Amjad, Kyunghyun Cho, Nizar Habash and Aaron Courville. (2016). *First Result on Arabic Neural Machine Translation.* ArXiv Preprints.

Baniata, Laith H., Seyoung Park and Seong-Bae Park. (2018). 'A neural machine translation model for Arabic dialects that utilizes multitask learning (MTL)'. *Computational Intelligence and Neuroscience*, 2018 (1): 1-10. https://doi.org/10.1155/2018/7534712

Baniata, Laith, Isaac Kojo Essel Ampomah and Seyoung Park. (2021). 'A transformer-based neural machine translation model for Arabic dialects that utilizes subword units'. *Sensors*, 21 (19): 6509.

Baziotis, Christos, Mikel Artetxe, James Cross and Shruti Bhosale. (2022). 'Multilingual machine translation with hyper-adapters'. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. United Arab Emirates: Association for Computational Linguistics, 1170–1185.

Bouamor, Houda, Nizar Habash and Kemal Oflazer. (2014). 'A multidialectal parallel corpus of Arabic'. *Proceedings of the Ninth Conference on Language Resources and Evaluation (LREC'14)*. Iceland: European Language Resources Association (ELRA), 1240–1245.

Bouamor, Houda, Nizar Habash, Mohammad Salameh, Wajdi Zaghouani, Owen Rambow, Dana Abdulrahim, Ossama Obeid, Salam Khalifa, Fadhl Eryani, Alexander Erdmann and Kemal Oflazer. (2018). 'The MADAR Arabic dialect corpus and lexicon'. *Proceedings of the Eleventh*

International Conference on Language Resources and Evaluation (LREC 2018). Japan: European Language Resources Association (ELRA), 3387–3396.

Chen, Song, Jennifer Tracey, Christopher Walker and Stephanie Strassel. (2019). BOLT Arabic Discussion Forum Parallel Training Data LDC2019T01. Web Download. Philadelphia: Linguistic Data Consortium (LDC).

Dabre, Raj, Chenhui Chu and Anoop Kunchukuttan. (2020). 'A survey of multilingual machine translation'. *ACM Computing Surveys*, 53 (5): 1–31.

Diab, Nesma. (2021). Out of the BLEU: 'An error analysis of statistical and neural machine translation of WikiHow articles from English into Arabic'. *CDELT Occasional Papers in the Development of English Education,* 75 (1): 181–211.

Farhan, Wael, Bashar Talafha, Analle Abuammar, Ruba Jaikat, Mahmoud Al-Ayyoub, Ahmed Bashir Tarakji and Anas Toma. (2020). 'Unsupervised dialectal neural machine translation'. *Information Processing and Management,* 57 (3): 102181.

Gadalla, Hassan, Hanaa Kilany, Howaida Arram, Ashraf Yacoub, Alaa El-Habashi, Amr Shalaby, Krisjanis Karins, Everett Rowson, Robert MacIntyre, Paul Kingsbury, David Graff, Cynthia McLemore. (1997). CALLHOME Egyptian Arabic Transcripts LDC97T19 [Corpus]. Web Download. Philadelphia: Linguistic Data Consortium (LDC).

Halliday, Michael A. K. and Christian M. I. M. Matthiessen. (2014). *Halliday's Introduction to Functional Grammar* (4th ed.). London and New York: Routledge.

Hamed, Injy, Nizar Habash, Slim Abdennadher, and Ngoc Thang Vu. (2022). 'ArzEn-ST: A three-way speech translation corpus for code-switched Egyptian Arabic–English'. *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, 119–130. United Arab Emirates Association for Computational Linguistics.

Harrat, Salima, Karima Meftouh, and Kamel Samili. (2019). 'Machine translation for Arabic dialects (survey)'. *Information Processing and Management*, 56 (2): 262–273.

Hassan, Hany, Mostafa Elaraby and Ahmed Y. Tawfik. (2017). 'Synthetic data for neural machine translation of spoken dialects'. *Proceedings of the 14th International Workshop on Spoken Language Translation*, 82–89. Japan: International Workshop on Spoken Language Translation.

Kchaou, Saméh, Rahma Boujelbane and Lamia Hadrich. (2023). 'Hybrid pipeline for building Arabic Tunisian dialect-standard Arabic neural machine translation model from scratch'. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22 (3): 1–21.

Kumar, Gaurav, Yuan Cao, Ryan Cotterell, Chris Callison-Burch, Daniel Povey and Sanjeev Khudanpur. (2014). Translations of the CALLHOME Egyptian Arabic corpus for conversational speech translation. *Proceedings of the 11th International Workshop on Spoken*

Language Translation: Papers. USA: Association for Computational Linguistics, 244–248.

Li, Xuansong, Stephen Grimes, and Stephanie Strassel. (2019). BOLT Egyptian Arabic–English Word Alignment -- SMS/Chat Training LDC2019T18. Web Download. Philadelphia: Linguistic Data Consortium (LDC).

Linguistic Data Consortium. (2002a). CALLHOME Egyptian Arabic Speech Supplement LDC2002S38 [Corpus]. Philadelphia: Linguistic Data Consortium (LDC).

Linguistic Data Consortium. (2002b). 1997 HUB5 Arabic Transcripts LDC2002T39 [Corpus]. Philadelphia: Linguistic Data Consortium (LDC).

Lison, Pierre and Jörg Tiedemann. (2016). 'OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles'. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'2016)*, 923–929. Slovenia: European Language Resources Association (ELRA).

Meftouh, Karima, Salima Harrat, Salma Jamoussi, Mourad Abbas, and Kamel Smaili. (2015). 'Machine translation experiments on PADIC: A parallel Arabic dialect corpus.' *Proceedings of the 29th Pacific Asian Conference on Language, Information and Computation*, 26–34. China.

Nagoudi, El Moatez Billah, AbdelRahim Elmadany and Muhammad Abdul-Mageed. (2021). 'Investigating code-mixed modern standard Arabic–Egyptian to English machine translation'. *Proceedings of the Fifth Workshop on Computational Approaches to Linguistics Code-Switching*, 56–64. Online: Association for Computational Linguistics.

Nagoudi, El Moatez Billah, AbdelRahim Elmadany and Muhammad Abdul-Mageed. (2022). 'TURJUMAN: A public toolkit for neural Arabic machine translation'. *Proceedings of the Open Source and Arabic Corpora and Processing Tools (OSCAT 2022) Workshop*, 1–11. France: European Language Resources Association (ELRA).

Richardson, Ashleigh and Janet Wiles. (2022). 'A systematic study reveals unexpected interactions in pre-trained neural machine translation'. *Proceedings of the Thirteenth Conference on Language Resources and Evaluation*, 1437–1443. France: European Language Resources Association (ELRA).

Saleh, Fahimeh, Wray Buntine, and Gholamreza Haffari. (2021). 'Collective wisdom: Improving low-resource neural machine translation using adaptive knowledge distillation'. *Proceedings of the 28th International Conference on Computational Linguistics*, 3413–3421. Spain: International Committee on Computational Linguistics.

Slim, Amelm, Ahlem Melouah, Usef Faghihi and Khouloud Sahib. (2022). 'Improving neural machine translation for low resource Algerian dialect by transductive transfer learning strategy'. *Arabian Journal for Science and Engineering*, 47: 10411–10418.

Sun, Mengtao, Hao Wang, Mark Pasquine and Ibrahim A. Hameed. (2021). 'Machine translation in low-resource languages by an adversarial neural network'. *Applied Science,* 11 (22): 10860.

Takezawa, Toshiyuki, Genichiro Kikui, Masahide Mizushima and Eiichiro Sumita. (2007). 'Multilingual spoken language corpus development for communication research'. *International Journal of Computational Linguistics and Chinese Language Processing: Special Issue on Invited Papers from ISCSLP 2006*, 12 (3): 303–324.

Tiedemann, Jörg. (2012). 'Parallel data, tools and interfaces in OPUS'. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 2214–2218. Turkey: European Language Resources Association (ELRA).

Tilde. (2023). Interactive BLEU Score Evaluator [Software]. https://www.tilde.com/products-and-services/machine-translation/features/interactive-bleu

United Nations Population Fund. (2023, April 23). *World Population Dashboard*. https://www.unfpa.org/data/world-population/EG

Pitman, Jeff. (2021, April 28). Google Translate: One billion installs, one billion stories. Google. https://blog.google/products/translate/one-billion-installs/ (Retrieved on May 13, 2023)

Wang, Zirui, Zihang Dai, Barnabas Poczos and Jaime Carbonell. (2019). 'Characterizing and avoiding negative transfer'. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 11293–11302 USA: Computer Vision Foundation. https://doi.org/10.1109/CVPR.2019.01155

Wang, Zirui, Zachary C. Lipton and Yulia Tsvetkov. (2020). 'On negative interference in multilingual models: Findings and a meta-learning treatment'. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4438–4450. Online: Association for Computational Linguistics.

Zakraoui, Jezia, Moutaz Saleh, Somaya Al-Maadeed and Jihad Mohamed Alja'am. (2021). 'Arabic machine translation: A survey with challenges and future directions'. *IEEE Access*, 9: 161445–161468.

Zbib, Rabih, Erika Malchiodi, Jacob Devlin, David Stallard, Spyros Matsoukas, Richard Schwartz, John Makhoul, Omar F. Zaidan and Chris Callison-Burch. (2012). 'Machine translation of Arabic dialects'. *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 49–59. Canada: Association for Computational Linguistics.

Zhang, Wen, Lingfei Deng, Lei Zhang and Dongrui Wu. (2023). 'A survey on negative transfer'. *IEEE/CAA Journal of Automatica Sinica*, 10 (2): 305–329.