Arabic Machine Translation: A survey of the latest trends and challenges

Mohamed Seghir Hadj Ameur *[a], Farid Meziane[b], Ahmed Guessoum[a]

[a]NLP, ML, and Applications Research Group (TALAA), Laboratory for Research in AI (LRIA), University of Science and Technology Houari Boumediene (USTHB), Algeria
[b]Data Science Research Centre, University of Derby, DE22 1GB, UK

Abstract

Given that Arabic is one of the most widely used languages in the world, the task of Arabic Machine Translation (MT) has recently received a great deal of attention from the research community. Indeed, the amount of research that has been devoted to this task has led to some important achievements and improvements. However, the current state of Arabic MT systems has not reached the quality achieved for some other languages. Thus, much research work is still needed to improve it. This survey paper introduces the Arabic language, its characteristics, and the challenges involved in its translation. It provides the reader with a full summary of the important research studies that have been accomplished with regard to Arabic MT along with the most important tools and resources that are available for building and testing new Arabic MT systems. Furthermore, the survey paper discusses the current state of Arabic MT and provides some insights into possible future research directions.

Keywords: Natural Language Processing, Machine Translation, Arabic Machine Translation, Arabic Language, Deep Learning

Contents

*Corresponding author
Email addresses: mhadjameur@usthb.dz (Mohamed Seghir Hadj Ameur *), f.meziane@derby.ac.uk (Farid Meziane), aguessoum@usthb.dz (Ahmed Guessoum)

1. Introduction

Machine Translation (MT) is a sub-field of Natural Language Processing (NLP) devoted to the development and enhancement of computer-based translation systems. The goal of an MT system is to automatically translate a given textual content from one language to another in a way that best preserves its meaning and style while ensuring that the produced translation output is as linguistically fluent as possible.

Machine translation has a very long history of creativity, research, and ambition. It has first been proposed by Warren Weaver in the late 1940s and, since then, its evolution has benefited from several factors such as the increase of computing power, the availability of large parallel corpora, and the rapid progress in the field of computer science and artificial intelligence [1] [1]. These factors have led to the emergence of Statistical Machine Translation (SMT) [4], the approach that has dominated the field of MT for the last two decades. Currently, another paradigm called Neural Machine Translation (NMT) [5] has also emerged and managed to push the boundaries of MT quality even further for many language pairs. These technologies have been used to build large-scale translation systems such as the well known Google [2] and Microsoft [3] translators which have provided the Internet users with high-quality instant translation services across several language pairs. A glimpse of the current translation quality that has been obtained on the task of MT between English and several Indo-European and Asian languages is provided in Figure 1 [4].

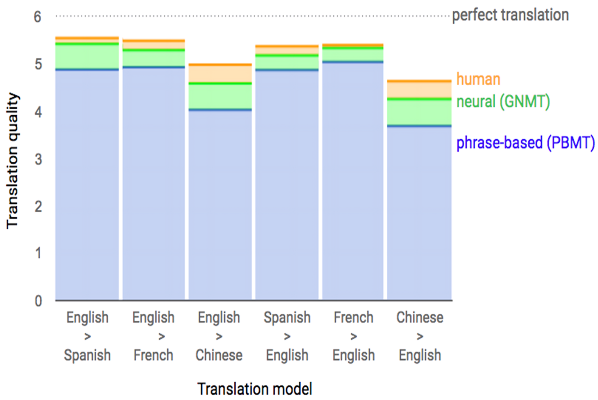

Figure 1: A comparison between Google's neural and statistical MT systems on the task of MT between English and several Indo-European and Asian languages [6]

Arabic is one of the five most spoken languages in the world with more than 300 million native speakers [5] and is one of the six official languages of the United Nation (UN) [6]. It is a Semitic language that is well known for its rich and complex morphology which is substantially different from that of Indo-European languages (such as English and French). The Arabic morphology added to other linguistic aspects has made the automatic translation from and to Arabic a lot more challenging. Though a great deal of improvement has been achieved due to the recent advances in data-driven translation paradigms (e.g. statistical and neural methods), these linguistic aspects are still causing many difficulties [7, 8].

In this paper, we provide an exhaustive survey of the important research studies that have been accomplished with regards to Arabic MT. Though an early survey was performed by Alqudsi et al. [9] a few years ago, we believe that our survey can be seen as an up-to-date survey as it covers many newly proposed research studies

[1] For a detailed history of machine translation, we point the reader to Hutchins [2], Hutchins and Somers [3], Hutchins [1].

[2] https://translate.google.com
[3] https://www.bing.com/translator/
[4] Figure 1 is taken from the Google AI Blog describing the paper of Wu et al. [6] https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html
[5] https://www.conservapedia.com/List_of_languages_by_number_of_speakers
[6] http://www.un.org/en/sections/about-un/official-languages/

that have not been covered previously, especially those dealing with neural machine translation and deep learning methods in general, which have revolutionized the field. We also propose a new classification of the research studies developed for Arabic MT based on the main problem that they address; this is also different from what has previously been proposed and also more practical for the reader. Moreover, this survey lists the most important tools and resources that are available for building and testing new Arabic MT systems, discusses the current state of the field, and provides some insights into possible future research directions.

The remainder of this paper is organized as follows. Section 2 specifies the scope and purpose of this survey. Section 3 introduces the Arabic language, its characteristics, and the difficulties faced during its translation to and from other languages. Section 4 gives a brief introduction to the current machine translation paradigms that are necessary for a better understanding of this survey. Then, Section 5 summarizes the research work that has been done with regard to Arabic MT. Section 6 lists the most important tools and resources that are available for building and testing new Arabic MT systems. Then, Section 7 provides a discussion of the current problems and limitations that are faced by the machine translation research community working on the Arabic language. Finally, section 8 gives a conclusion to this survey and provides some possible future research directions.

2. Scope of the Study

This survey paper aims to present the machine translation studies that have been developed regarding Arabic MT in a categorized and easy-to-read manner. It considers only the translation studies related to the formal Modern Standard Arabic (MSA) language. Thus, the translation studies for informal Arabic (regional dialects) will not be covered in this survey. This survey is primarily intended for researchers and scholars who want to have an up-to-date overview of what has been accomplished so far in the field of Arabic MT. It also provides a quick overview of the Arabic language characteristics and translation difficulties.

This survey adopts the Buckwalter transliteration scheme[7] to transcribe Arabic words using Latin characters.

[7] Buckwalter uses a simple one-to-one mapping between the Arabic and Latin character sets `https://en.wikipedia.org/wiki/Buckwalter_transliteration`

3. The Arabic Language: Characteristics and Translation Challenges

In recent years, the amount of research work that has been devoted to Arabic natural language processing has considerably increased. Due to its rich and complex morphology, there has been an increasing demand for highly sophisticated Arabic NLP tools that can satisfy its growing needs in several domains and applications.

In this section, we briefly introduce the Arabic language along with the difficulties that are involved in its translation. Given that the main concern of the research community is the task of translation between Arabic and English, this section mainly focuses on the translation difficulties concerning these two languages. We note that this introduction about the Arabic language and its translation challenges merely scratches the surface in this area, thus we refer the reader to Habash [10], Ryding [11], and Farghaly and Shaalan [12] for more detailed overviews about the Arabic language and its characteristics.

3.1. Arabic Language Characteristics

In the following subsections, we will highlight some orthographic, morphological, and syntactic aspects of the formal Arabic language (MSA). At the same time, we will try to compare the linguistic aspects of the language to those of some other languages such as English.

3.1.1. Arabic Orthography

Orthography studies the spelling system of a language. This generally involves a set of symbols that are used to write the language and a set of rules that indicate their usage. Arabic contains 28 letters which make its orthographic complexity comparable to that of English, and much simpler than the Chinese orthography which contains around 10,000 logographic characters [8] [10]. The shape of a single Arabic letter may change slightly depending on its position within the Arabic word (beginning, middle, or at the end). Unlike Latin languages, Arabic texts are written from right to left and do not use special letters to represent vowels. Instead, the Arabic writing system uses diacritics which are small marks that can be added above or below the letters. The presence of these diacritical marks is known as "Vocalization". When the vocalization is present, it adds more information about the correct pronunciation and meanings of the words which solves many lexical and semantic

[8] A Logogram is a single character that can represent a morpheme, a word, or even an entire phrase [13, 14].

ambiguities [15]. The most common diacritical marks (also known as short vowels) in the Arabic language are Fathah as in أَ (makes an "a" sound as in "Adam"), Dammah as in أُ (makes an "oo" sound as in "took"), and Kasrah as in إِ (makes an "i" sound as in "in").

3.1.2. Arabic Morphology

Morphology studies the internal structure of words and emphasizes the way morphemes [9] fuse with each other to form new words and express new meanings. The Arabic morphology involves two main operations: inflection and derivation [11].

Inflection is the mechanism that allows the creation of new words (inflected words) from other words by adding inflectional morphemes that express specific grammatical properties (e.g., gender, person, number, aspect, mood, time, etc.). This process generally retains both the meaning and the syntactic category of the base word. Arabic is known to be a highly inflected language. Its verbal inflection usually follows very regular patterns with hardly any exceptions [10]. It has two gender values: masculine (ex. يَسْمَعُ, yasomaEu, he listens) and feminine (ex. تَسْمَعُ, tasomaEu, she listens); three number values: singular (ex. يَسْمَعُ, yasomaEu, he listens), dual (ex. يَسْمَعَانِ, yasomaEaAni, they listen) and plural (ex. يَسْمَعُونَ, yasomaEuwna, they listen); three tenses: the past (ex. سَمِعَ, samiEa, he listened), the present (ex. يَسْمَعُ, yasomaEu, he is listing), the future (ex. سَيَسْمَعُ, sayasomaEu, he will listen); and two voices: passive (ex. سُمِعَ, sumiEa, it was listened) and active (ex. سَمِعَ, samiEa, he listened). Arabic nominal morphology is a lot more complex than the verbal one [10]; it includes gender, number, state, and case inflections. It has two gender values: masculine (ex. سَرِيعٌ, sariyEN, fast) and feminine (ex. سَرِيعَةٌ, sariyEapN, fast); three number values: singular (ex. مُعَلِمٌ, muEalimN, teacher), dual (ex. مُعَلِمَانِ, muEalimaAni, teachers) and plural (ex. مُعَلِمُونَ, muEalimuwna, teachers); three states: definite (ex. الرَجُلُ, Alrajulu, the man), indefinite (ex. رَجُلٌ, rajulN, man), and construct (ex. رَجُلُ الكَهْفِ, rajulu Alkahofi, caveman); and three case values: nominative (ex. المَدْرَسَةُ, Almadorasapu, the school), accusative (ex. المَدْرَسَةَ, Almadorasapa, the school), genitive (ex. المَدْرَسَةِ, Almadorasapi, the school). Unlike Arabic, English is not a highly inflected language. Its verbs inflect mainly for number (ex. the lion **eats**), the past tense (ex. helped), and the present tense (ex. looks). On the other hand, its nouns inflect only for the plural (ex. lions) and the possession (ex. John's).

Derivation is the mechanism that allows the creation of new words (inflected words) from other words by inserting one or multiple affixes. This process generally modifies not only the core meaning of the word but also its syntactic category. The English language supports only very basic derivations. These derivations are generally obtained by adding prefixes (ex. do/**un**do, do/**re**do, way/**sub**way), suffixes (ex. reason/reason**able**, way/ way**ward**) or both (ex. reason/**un**reason**able**). Arabic derivational morphology is different from that of Indo-European languages which rely mostly on the concatenation of stems and affixes. Indeed, Arabic, among other Semitic languages, is based on a root-pattern system which uses two important components to compose words: roots and patterns. A Root also known in Arabic as الجِذر ("Alji*r") holds the basic semantic information (or an abstract meaning) that is shared between all the words that can be derived from it. We note that the Arabic roots are categorized based on the number of their unvocalized letters into triliteral (containing three letters), quadriliteral (containing four letters), and quintiliteral (containing five letters) roots [10]. Patterns are abstract templates that take the basic root word and allow the creation of derivative words that have a certain syllabic structure and hold syntactic and semantic information. Table 1 presents a few derivations of the root "ك ت ب".

Table 1: An example illustrating some derivations of the root "ك ت ب"

Arabic Derivation	Buckwalter	Meaning
كِتَابَةٌ	kitaAbapN	writing
كَاتِبٌ	kaAtibN	a writer
مَكْتُوبٌ	makotuwbN	written
مَكْتَبَةٌ	makotabapN	a library
كِتَابٌ	kitaAbN	a book

The derivations of an Arabic root can also be com-

[9] A Morpheme is the smallest part of a word that has a significant meaning in a specific language.

bined with other prefixes and suffixes resulting in very morphologically rich words that can hold a considerable amount of information. Indeed, a single Arabic word can have the meaning of multiple words or even a whole sentence in other languages. For instance, the Arabic word "أفنلزمكموها" ("¿fnlzmkmwhA") has the meaning of the whole sentence "shall we then force you to do/accept it".

3.1.3. Arabic Syntax

Syntax studies the mechanisms responsible for arranging words within a language to create meaningful phrases and sentences. When dealing with a morphologically rich language such as Arabic these mechanisms become heavily related to the morphological level. Thus, several syntactic aspects are not expressed uniquely via word order but also through morphology [10].

The Arabic language admits two types of sentences: verbal and nominal. A verbal sentence in Arabic starts with a verb. In its most basic form, it must contain a verb and a subject within a Verb-Subject word order. For example, the sentence "خرج الولد" (the boy went out) is a simple Verb-Subject sentence in which the verb is "خرج" (went out) and its subject is "الولد" (the boy). The object can also be present after the subject as in the sentence "كتب التلميذ الدرس" (the pupil wrote the lesson) which is a Verb-Subject-Object sentence in which the verb is "كتب" (wrote), its subject is "التلميذ" (the pupil), and the object is "الدرس" (the lesson). This is quite different from the English sentence structure which generally follows a Subject-Verb-Object word order.

A nominal sentence in Arabic does not require a verb. This kind of sentence is not present in English given that each sentence in the English language requires at least one verb. Nominal sentences in Arabic start with a noun and have two parts: a subject and a predicate. The subject (topic) needs to be a noun and the predicate can be a noun, an adjective, a pseudo-sentence [10], or a verbal sentence. For example, the sentence "الشمس مشرقة" (the sun is shining) is a simple Noun-Adjective nominal sentence resulting from the concatenation of two words: "الشمس" (the sun) and "مشرقة" (shining).

An important feature of Arabic syntax is the fairly loose word order. For instance, the different orderings of the three Arabic words أَكَلَ (ate), الولدُ (the boy), and التفاحةَ (the Apple) convey slightly different meanings[11]:

1. "أَكَلَ الولدُ التفاحةَ" in a Verb-Subject-Object order has the meaning of "the boy ate the Apple".
2. "الولدُ أَكَلَ التفاحةَ" in a Subject-Verb-Object order has the meaning of "(it is) the boy (who) ate the Apple".
3. "التفاحةَ أَكَلَ الولدُ" in an Object-Verb-Subject order has the meaning of "(it is) the Apple (that) the boy ate".
4. "أَكَلَ التفاحةَ الولدُ" in a Verb-Object-Subject order has the meaning of "the boy ate the apple".

Even Though all the above possibilities are accepted in the Arabic language, the first one is the most common. This kind of flexibility is not found in the English language.

3.2. Arabic Machine Translation Challenges

Besides the above morphological and syntactic characteristics of the Arabic language that complicate the task of its translation, many other known problems pose serious challenges that are heavily studied by the Arabic MT community. We will briefly highlight the most important ones in the following subsections [12].

3.2.1. Arabic Vocalization

Arabic texts can be fully, partially, or not vocalized at all which causes many challenges to Arabic natural language processing applications [18, 15]. A common method to address this problem is to perform a preprocessing step that removes all the diacritical marks and produces a fully unvocalized Arabic text.

Using unvocalized Arabic texts reduces the number of word forms (vocabulary size) which is often a positive thing for MT [19]. However, the absence of vocalization may cause serious ambiguities, especially when dealing with short texts with a limited context. Indeed, a single unvocalized Arabic word can have multiple meanings. For instance, depending on its vocalization, the Arabic word "ولد" (wld) can have all the meanings presented in Table 2.

Even in the presence of some contextual words, the unvocalized Arabic words can still be ambiguous, or at

[10]A pseudo-sentence in Arabic ("شبه الجملة") refers simply to a phrase that contains a preposition followed by a noun.

[11]This Arabic word ordering example is inspired from the one given in [9].

[12]For a detailed overview of Arabic MT difficulties we refer the reader to [7, 16, 17].

Table 2: An example illustrating multiple meanings that are driven from the same unvocalized Arabic word "ولد" (wld)

Word	Buckwalter	Meaning
وُلِدَ	wulida	he was born
وَلَدَ	walada	he gave birth to
وَلَدُ	waladu	the son of
وَلَدٌ	waladN	a boy
وَلَّدَ	wala~da	helped someone else give birth

least demand a considerable automatic natural language analysis to disambiguate them. For instance, if we consider the following simple Arabic sentence:

"ولد ذلك الشخص الذي رأيناه فار من العدالة"
(the son of that man that we have seen, is fleeing from justice)

The word "ولد" in this context means "the son of" ("وَلَدُ"), yet it is quite difficult for MT systems to figure this out, thus, they often fail to translate it. For example, the Google Translation Service [13] translates it as "that person whom we saw was born out of justice" which is wrong, as it assumes that "ولد" has the meaning of "وُلِدَ" (he was born), instead of "وَلَدُ" (the son of).

3.2.2. Arabic Words Polysemy

A polyseme refers to a word (or a phrase) that has multiple senses (meanings) [20]. The task of identifying the correct sense of a given word that has multiple meanings (an ambiguous word) in a given sentence (or text) is known as Word Sense Disambiguation (WSD). As we illustrated in the previous section, the vocalization of a word can help solve some lexical and semantical ambiguities in Arabic, yet even if the word is fully vocalized, it can still have several meanings depending on how it is used. For instance, the word "عَيْنُ" ("Eayonu") in Arabic has multiple meanings in English as shown in Table 3.

As shown in Table 3 the meaning of a single Arabic word can change depending on its context. For MT, this polysemy problem is not limited to Arabic but to both the source and the target languages that are involved. Indeed, if we consider the task of translation between English and Arabic then the problem of polysemy may

Table 3: An example illustrating a vocalized Arabic word that has several possible meanings

Arabic	English
عَيْنُ الماء	water source
عَيْنُ الرجل	man's eye
عَيْنُ المكان	the place itself

also come from the English language. For example, the English word "take" has several meanings in Arabic as shown in Table 4, where its meaning changes when associated with different prepositions (e.g. out, off, up, down, back, on, in, by, etc.).

Table 4: An example illustrating several meanings that the word "take" can express depending on its context

English	Arabic
take me to a place	خذني إلى مكان
take them down	إقضِ عليهم
take care of it	إهتم بالأمر
take a long time to load	يستغرق وقتًا طويلاً للتحميل
take off from the airport	تقلع من المطار
take back what you said	تراجع عن كلامك

3.2.3. Arabic Multiword Expressions

A Multiword Expression (MWE) refers to a collocation of two or more words that appear together but whose meaning is not deducible or is only partially deducible from the semantic meanings of their constituents [21, 22]. Arabic MWEs are mainly idiomatic expressions that are frequently used by Arabic speakers in different contexts and whose translation is not evident. Some examples of Arabic idioms are given in Table 5.

MWEs have several aspects that make them difficult to translate [23], from which we cite the two main ones:

1. Non-compositionality: The overall expression of semantics is not related to individual pieces or words. This is the case, for example, if we consider the Arabic expression "طويل اللسان" which has the literal meaning of "long tongue" in English, where the individual words "طويل" and "اللسان" translate to "long" and "the tongue", respectively. However, this expression means a rude or indecent person.

[13] https://translate.google.com

Table 5: Some examples of Arabic idioms

Arabic idioms	Literal words meanings	Idiomatic meaning
ما باليد حيلة	no trick in hand	cannot be helped
بشق الأنفس	by cracking the souls	with enormous difficulty
خفيف الظل	with/having a light shadow	funny/pleasant

2. MWE and non-MWE expressions ambiguity: It is hard to determine if the MWE is intended for its literal or idiomatic meaning. For instance, if we consider the Arabic phrase "حيوان طويل اللسان" (an animal that has a long tongue), the MWE "طويل اللسان" here has its literal meaning, not its idiomatic one.

3.2.4. *Arabic Named Entities*

Named entities (NEs) refer to all the entities that can be referenced by using proper names [24], such as persons, locations, organizations, quantities, values, etc. The task of identification, extraction, and classification of these entities is known as Named Entity Recognition (NER). Depending on the application, this task generally considers four main categories: person names (PER) such as "أحمد" (Ahmed), locations (LOC) such as "الجزائر" (Algeria), organizations (ORG) such as "سامسونج" (Samsung), and miscellaneous (MISC) which includes all remaining types of entities [25]. The task of NER is known to be much more difficult when performed on the Arabic language in comparison to most of the Latin languages due to many specific linguistic aspects that characterize Arabic. The difficulty is mainly due to the lack of capitalization, the rich lexical variations, and the lack of uniformity in writing Arabic named entities [25]. In the field of MT, a proper handling of named entities is crucial to produce a reliable translation result. Indeed, named entities need to be addressed carefully as two possible treatments can be adopted depending on either a meaning-based translation or a phoneme-based transliteration [25, 26]:

1. Arabic meaning-based Translation: In this case, entities are translated according to their meaning. For example "الولايات المتحدة الأمريكية" (AlwlAyAt AlmtHdp Al¿mrykyp) gets translated to "The United States of America".
2. Arabic Transliteration: Here entities (e.g. personal names) are transliterated. For example "نيويورك" (nywywrk) gets transliterated to "New York".

4. Machine Translation Paradigms

In this survey, we classify the existing machine translation approaches based on the sources of information required for their construction as suggested by Costa-Jussa and Fonollosa [27]. These sources can be either rules (i.e. rule-based methods) or data (i.e. data-driven methods). Combining these two sources will lead to a third category of methods known as hybrid MT approaches. Figure 2 presents the main translation paradigms that can be found under each of these categories.

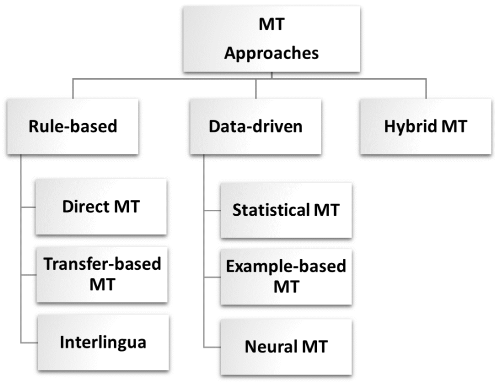

Figure 2: Hierarchical categorization of the main machine translation approaches [27]

4.1. *Rule-based Machine Translation*

Rule-based Machine Translation (RBMT) is the first class of methods that have been used in the field of MT [3]. These kinds of methods incorporate linguistic information (e.g. handcrafted rules) about the source and target languages that vary from low-level morphological information to high-level syntactic and semantic information. Indeed, rule-based MT methods can be divided into three classes: (1) the direct MT methods which attempt to model the relation between the source and target languages by relying only on morphology-level information, (2) transfer methods which involve some syntax level knowledge, and (3) the interlingua methods which attempt to model the semantic level. The linguistic level of analysis involved in each class of methods increases gradually from the lowest direct MT level to the highest interlingua semantic-based translation; these methods are generally represented via the "Vauquois triangle" shown in Figure 3 where the direct MT methods are placed at the bottom of the triangle given that they involve almost no linguistic analysis. As we go higher in the triangle the level of analysis which is required increases until we reach the Interlingua MT which involves the highest level of linguistic analysis. In the following, we will briefly introduce the direct, transfer, and interlingual rule-based translation methods.

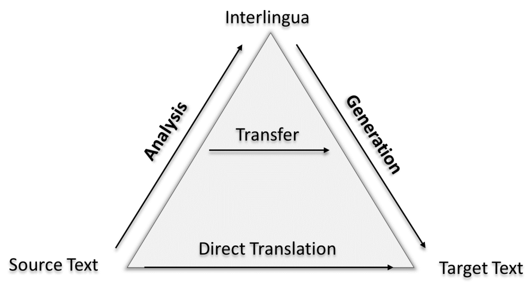

Figure 3: The Vauquois triangle

4.1.1. Direct Machine Translation

The direct MT approach directly translates a source text to a target text without involving any linguistic analysis that goes beyond the morphological level [3]. Figure 4 shows the overall translation phases involved in the direct MT approach. First, the source text is morphologically analyzed, then, the translation process is carried out word by word by using a large bilingual dictionary that maps each source word to a target word. Each entry (word) in the bilingual dictionary is generally associated with some simple rules that specify its different translation scenarios.

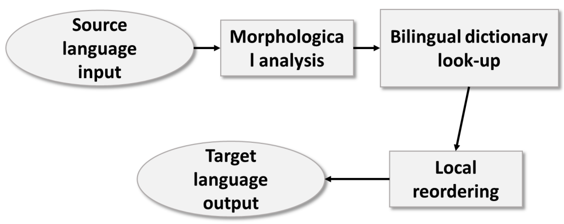

Figure 4: Direct machine translation phases [3].

An example of such a procedure is given in Figure 5. It shows a set of handcrafted rules that have been written by experts to cover the different translation scenarios of the two words "much" and "many" from English to Russian [24]. After each word is translated, some simple reordering rules are applied (e.g. rules for moving adjectives after nouns) to arrange the target words in a correct syntactic order.

```
function DIRECT_TRANSLATE_MUCH/MANY(word) returns Russian translation

if preceding word is how return skol'ko
else if preceding word is as return stol'ko zhe
else if word is much
   if preceding word is very return nil
   else if following word is a noun return mnogo
else /* word is many */
   if preceding word is a preposition and following word is a noun return mnogii
   else return mnogo
```

Figure 5: A simple procedure that guides the translation process of "much" and "many" from English to Russian under the direct MT method [24]

4.1.2. Transfer-based Machine Translation

Instead of considering the translation as a direct mapping process, transfer-based MT methods [3, 24] use an intermediate structural representation to capture the different linguistic aspects of the input sentence. Then, the translation output is generated from that representation using a set of sophisticated transfer rules.

The functioning mechanism of this approach (Figure 6) can be summarized in three consecutive steps: analysis, transfer, and generation. First, an analysis step is performed in which the source sentence is analyzed both morphologically and syntactically to create an internal representation that captures its syntactic aspects (e.g. a parse tree of the source sentence). Then, a set of linguistic rules are used to transform the structural representation of the input sentence into the target one. The morphological and syntactical differences that exist between the source and target languages are also handled at this stage via these transfer rules. Finally, the output translation is generated from the resulting target representation by using large bilingual dictionaries.

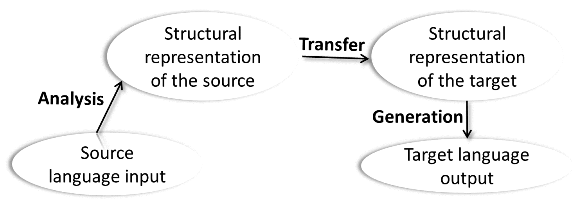

Figure 6: The steps involved in a transfer-based MT system

Figure 7 illustrates the steps needed to translate a simple English sentence into French using the transfer-based translation approach. It shows that the English sentence structural representation (parse tree) is transformed into a French parse tree by applying a simple Noun-Adjective syntactic reordering rule and using a bilingual English-to-French dictionary.

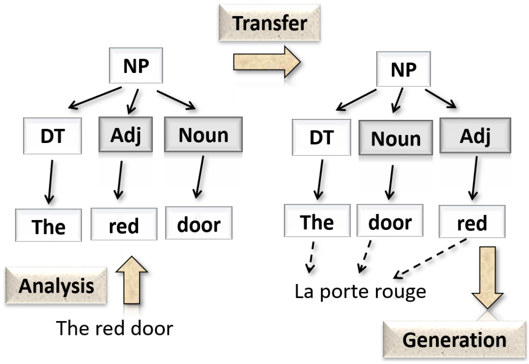

Figure 7: Example of Transfer-based approach between English and French.

The downside of this approach is the high cost of building and maintaining a consistent set of transfer rules that transform the structural representation of the source language into the target one.

4.1.3. Interlingual-based Machine Translation

One main problem of transfer-based MT is that it requires the development of a separate set of transfer rules to handle the translation between each pair of languages. Instead of building a specific language-related representation of the source sentence and then transform it to the target one, the interlingual-based MT approach uses a pivot representation called "Interlingua", that is independent of any natural language [3]. This representation holds the meaning of a given sentence in a purely abstractive way regardless of its source language. This means that the transfer component of the transfer-based MT (shown in Figure 6) is no longer needed leaving the interlingual-based MT with only two steps: analysis and generation as shown in Figure 8. The analysis step involves all the necessary morphological, syntactical, and semantic linguistic analysis needed to transform the source language text into an abstractive representation of meaning "Interlingua". Then, the generation step performs the reverse processing in which the target output is generated from the interlingual representation.

This approach has the advantage of being extremely suited for multilingual translation systems given that all the language pairs will share the same abstract representation (no transfer component is needed). However, it suffers from the fact that in practice building

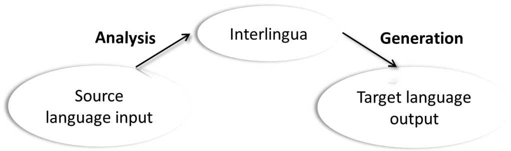

Figure 8: Interlingua-based machine translation phases

such a universal language-independent representation is extremely hard. Consequently, such methods are only used in very specific domains and still cost a huge amount of human effort [24].

4.2. Data-driven Machine Translation

Unlike the rule-based methods which rely on human knowledge (e.g. handcrafted rules), data-driven methods use sophisticated algorithms and mathematical models to automatically learn the translation process from data. This class of approaches relies heavily on large bilingual aligned corpora which are currently available across multiple languages. The availability of such corpora and the fact that this class of methods does not require any human knowledge have allowed it to dominate the field of MT for the past couple of decades. Data-driven MT includes example-based MT (EBMT), statistical MT, and neural MT approaches.

4.2.1. Example-based Machine Translation

Translating by example was first proposed by Nagao [28] and the idea behind it is to translate by analogy. Indeed, it follows the intuition that people do not translate using deep linguistic analysis; instead, they split the text into smaller segments and translate them separately, then they combine those smaller segments to produce the final translation. Thus, to translate a new input text, this approach attempts to adapt the previously translated sentences to produce its translation instead of trying to translate it from scratch. This approach uses a large bilingual corpus that contains a set of parallel texts (segments) as its knowledge base. The phases involved in this method are shown in Figure 9.

As illustrated in Figure 9, this approach involves three modules:

1. The matching module attempts to decompose the input sentence into smaller fragments.
2. The identification module (alignment module) tries to align each segment into its best matching translation using the example database.
3. The recombination module tries to combine the translated fragments (phrases) to form the translation output.

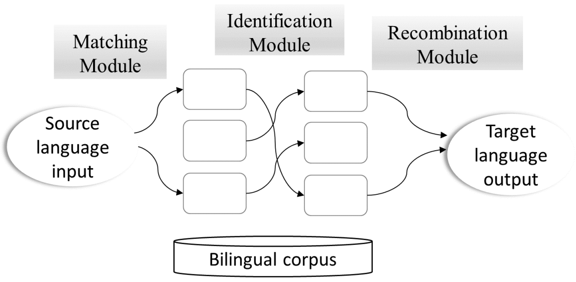

Figure 9: The steps involved in Example-based MT

Even though this approach does not require any handcrafted linguistic rules, its performance is heavily influenced by the quality of the example database.

4.2.2. Statistical Machine Translation

The SMT paradigm was proposed in the early 90s by a group of researchers working at the IBM Watson Research Center [29, 30, 31] and has evolved to become one of the leading approaches in the field of MT. Unlike those based on the rule-based methods, SMT systems are built automatically using monolingual and parallel corpora.

Given a source sentence f along with its translation e, the problem of SMT can be formulated as follows:

$$\hat{e} = argmax_e p(e|f) \tag{1}$$

The goal is to find the best translation ê that maximizes $p(e|f)$, the probability of e being the translation of f [29, 30]. The noisy channel model [30, 4] is used to decompose $p(e|f)$ into a translation model $p(f|e)$ and a language model $p(e)$.

$$\hat{e} = argmax_e p(f|e) * p(e) \tag{2}$$

The components of a statistical machine translation system are provided in Figure 10.

The SMT components resulting from the noisy channel formulation are as follows:

- A language model for computing $p(e)$ which ensures the fluency of the generated target output. The language model is built using a monolingual target corpus.
- A translation model for computing $p(f|e)$ which ensures the accuracy of the translation between the source and the target languages. The translation model is built using a parallel source-target corpus.
- A decoder, which is used to find the most probable translation output ê for the input sentence f from the space of all possible translations of f.

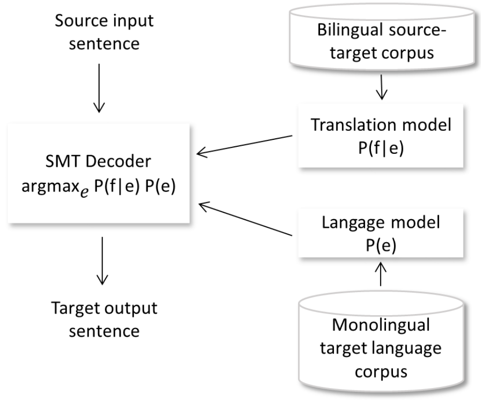

Figure 10: The components of a statistical machine translation system

4.2.3. Neural Machine Translation

NMT [32, 5] is a recently proposed approach for tackling the MT task. Unlike the traditional SMT approach which involves several components that are tuned separately, NMT uses a single large neural network that is tuned at once to increase the translation quality [5].

Given a source sentence $X = (x_1, x_2, ..., x_d)$ and a target sentence $Y = (y_1, y_2, ..., y_{d'})$ where each x_t and y_t represent the source and target words at time-step t, d and d' representing the maximum source and target sentence lengths respectively, the NMT model estimates the conditional probability of generating the target sentence Y given the source sentence X as $P(Y = (y_1, y_2, ..., y_{d'})|X = (x_1, x_2, ..., x_d))$.

Its architecture (Figure 11) involves two components: an encoder and a decoder.

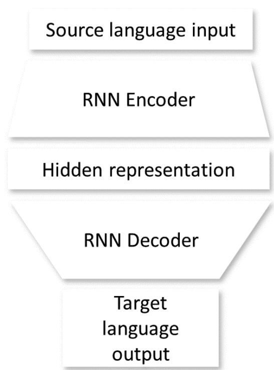

Figure 11: The components of a neural machine translation system [32]

The encoder is usually implemented as a bidirectional Recurrent Neural Network (bRNN) [33] that reads the input sentence word by word from left-to-right and right-to-left (reversed direction). The decoder is generally a unidirectional recurrent neural network that uses an attention mechanism that allows it to pay attention to the most relevant information present in the encoder hidden representation. In practice, two specific types of RNNs are used to implement the encoder and the decoder components which are the Long Short-term Memory (LSTM) [34] and the Gated Recurrent Unit (GRU) [35]. The NMT model is trained to maximize the probability of generating the target sentence given the source sentence in an end-to-end learning manner via the back-propagation algorithm based on a substantially large parallel corpus.

4.3. Hybrid Machine Translation

Hybrid machine translation is a class of methods that attempt to combine the aspects of several translation approaches into a single translation system. The goal of these methods is to benefit from the advantages of each approach and to produce a better translation. There are two main hybridization methods [27], hybrid rule-based MT, and hybrid data-driven MT.

4.3.1. Hybrid rule-based MT

These hybridization methods attempt to use data information (extracted via data-based methods) into a pre-existing rule-based translation system. For example, word-based alignment methods (e.g. IBM 1-5 or HMM alignment methods [36]) have been used by Habash et al. [37] to enhance the bilingual dictionaries of a rule-based system by adding new entries (e.g. new words or phrases) extracted automatically from a bilingual corpus.

4.3.2. Hybrid data-driven MT

These approaches attempt to either include rules into an existing data-based MT system (e.g. statistical or example-based systems) or to combine several data-driven methods into a single MT system [27].

- Enhancing a system using rules: the idea here is to use rules to enhance a preexisting data-driven system via a preprocessing or a post-processing task. For example, handcrafted rules can be used to change the syntactic order of the source language to match the target one [38].

- Enhancing a system using combination: the idea here is to combine several data-driven approaches into a single MT system to boost their performance. For instance, Groves and Way [39] proposed a hybrid system that combines subsentential alignments from both a phrase-based and an example-based MT systems and managed to outperform both of these individual systems.

5. Arabic Machine Translation: Research work

The main focus of the Arabic MT community has been devoted to translating Arabic into English. Less interest has been given to the English-to-Arabic translation direction and even fewer research studies have been devoted to translating Arabic to other languages than English. The research studies developed in the field of Arabic MT have been mainly devoted to statistical and neural MT; other translation approaches have received less attention. In this section, we attempt to classify the Arabic MT research work based on the main translation approaches that have been followed (e.g. rule-based, statistical, etc.). Given that most of these research studies fall under the SMT and NMT paradigms, we have considered further categorizing them based on the main contribution of each research work. Our detailed classification of the Arabic MT studies is presented in Figure 12.

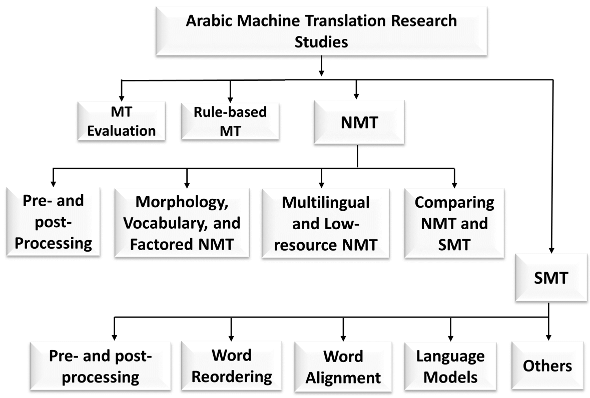

Figure 12: Classification of Arabic MT research work

Table 6 presents the main contributions that have been made in the field of Arabic MT in a manner that follows our proposed classification (Fig. 12). The studies will be detailed in the remainder of this section.

5.1. Research Studies on Arabic Statistical Machine Translation

Research work on SMT systems have primarily focused on one of the following aspects:

1. Improving the quality of statistical MT systems by performing a basic morphological pre- or/and post-processing of the Arabic language.

Table 6: The main contributions in Arabic MT

Main Research Area	Sub-research Area	Main Research Studies
SMT	Pre- and Post-processing	**Arabic-to-English:** Lee [40], Habash and Sadat [8], Sadat and Habash [41], Diab et al. [19], Hasan et al. [42] **English-to-Arabic:** Al-Haj and Lavie [43], Badr et al. [44], El Kholy and Habash [45]
	Word Reordering	**Arabic-to-English:** Chen et al. [46], Habash [47], Carpuat et al. [48], Carpuat et al. [49], Bisazza et al. [50] **English-to-Arabic:** Elming and Habash [51], Elming [52], Badr et al. [53], Hadj Ameur et al. [54] **Arabic-Others:** Sadat and Mohamed [55], Mohamed and Sadat [56], Alqudsi et al. [57]
	Word Alignment	**Arabic-to-English:** Ittycheriah and Roukos [58], Fossum et al. [59], Hermjakob [60], Gao et al. [61], Riesa and Marcu [62], Khemakhem et al. [63] **English-to-Arabic:** Ellouze et al. [64], Berrichi and Mazroui [65]
	Language Models	**Arabic-to-English:** Brants et al. [66], Carter and Monz [67], Niehues et al. [68] **English-to-Arabic:** Khemakhem et al. [69]
	Other Arabic SMT Studies	**Arabic-to-English:** Habash [70], Marton et al. [71] **English-to-Arabic:** Toutanova et al. [72]
NMT	Pre- and Post-processing	**Between Arabic & English:** Sajjad et al. [73], Oudah et al. [74], Hadj Ameur et al. [75] **Arabic-Others:** Aqlan et al. [76]
	Morphology, Vocabulary, and Factored NMT	**Between Arabic & English:** Ding et al. [77], Ataman et al. [78], Ataman et al. [79], Liu et al. [80], Shapiro and Duh [81] **Arabic-Others:** Belinkov et al. [82], Ataman and Federico [83], García-Martínez et al. [84]
	Multilingual & Low-resource	**Between Arabic & English:** Nishimura et al. [85], Tan et al. [86], Liu et al. [87], Aharoni et al. [88] **Arabic-Others:** Almansor and Al-Ani [89], Ji et al. [90]
	Comparing NMT & SMT	**Between Arabic & English:** Almahairi et al. [91], Junczys-Dowmunt et al. [92] **Arabic-Others:** Belinkov and Glass [93]
Rule-based MT	-	**Arabic-to-English:** Salem et al. [94] Shirko et al. [95] **English-to-Arabic:** Soudi et al. [96], Shaalan et al. [97], Al Dam and Guessoum [98]
MT Evaluation	-	**Arabic-to-English:** Hadla et al. [99] **English-to-Arabic:** Guessoum and Zantout [100], Guessoum and Zantout [101], Adly and Al Ansary [102], Al-Rukban and Saudagar [103], Guzmán et al. [104], El Marouani et al. [105]

2. Using a syntactic reordering process to decrease the syntactic gap between the source and the target languages.
3. Improving the word alignment quality as a means of increasing the overall SMT translation performance.
4. Incorporating rich linguistic features via additional language models.

5.1.1. Morphological Pre- and Post-Processing

The Arabic language is known to have a very rich and complex morphology. To tackle this, the majority of the research studies on Arabic SMT have primarily focused on performing morphological pre- and/or post-processing of the language before the translation task.

Arabic-to-English. Lee [40] proposed a model that attempts to establish a syntactic symmetry between Arabic and English languages which have a very asymmetrical morphology. Her method aligns a word-segmented and POS-tagged Arabic corpus to a symbol-tokenized English corpus using IBM word alignment model 1 [36]. She tested her proposal on an Arabic-to-English phrase-based SMT baseline with parallel corpora of different sizes and reported that the morphological preprocessing is helpful only when dealing with a small corpus. As the size of the training data increased, the benefit from performing such a preprocessing decreased. Habash and Sadat [8] and Sadat and Habash [41] studied the impact of several Arabic morphological preprocessing schemes on the task of Arabic-to-English phrase-based statistical MT. They found that splitting off only the conjunction and particles gives the best performance when using a large amount of training data (they called this scheme D2). However, if the amount of training data is limited, then more sophisticated morphological analysis and disambiguation are needed to achieve better performance. They concluded that an appropriate preprocessing produces a significant increase in the BLEU score, especially when dealing with test data that is not fairly similar to the training one. Diab et al. [19] investigated the effect of using different diacritization schemes that range from partial to full diacritization. They tested these schemes on an Arabic-to-English SMT baseline which involves no diacritization and found that the partial diacritization schemes do not have any significant impact on the baseline performance. They also found that a full diacritization scheme produces significantly worse results than the SMT baseline. Hasan et al. [42] investigated the effectiveness of using a rule-based FST segmenter and an SVM-based statistical segmenter (from the MADA toolkit[14]) on the

[14]MADA [106] is an Arabic preprocessing toolkit that handles tokenization, diacritization, morphological disambiguation, POS tagging, stemming, and lemmatization.

source Arabic language. They tested their proposal on an Arabic-to-English SMT task and found that the rule-based segmenter performs poorly when dealing with large translation tasks and that the use of the MADA statistical tool, has yielded better performance but at the cost of a much slower speed.

English-to-Arabic. Al-Haj and Lavie [43] also investigated the impact of using Arabic morphological preprocessing. They tested several segmentation schemes ranging from fully unsegmented to fully segmented Arabic forms on an English-to-Arabic phrase-based SMT baseline and found that the main gain comes from splitting up the pronominal enclitics in their investigated schemes. Badr et al. [44] and El Kholy and Habash [45] tested the impact of pre- and post-processing the Arabic target language in English-to-Arabic SMT. The Arabic texts were first decomposed morphologically in a preprocessing phase, then recombined in a post-processing step via several recombination techniques. The authors showed that morphological processing leads to a significant improvement, especially when using a small parallel training corpus.

5.1.2. Syntactic Word Reordering

Some works have used a specific kind of preprocessing known as "syntactic reordering" which aims at decreasing the syntactic differences between the source and the target languages by using syntactic reordering rules.

Arabic-to-English. Chen et al. [46] proposed a reordering method that automatically extracts reordering rules from a parallel corpus and uses them to address the reordering phenomena in phrase-based SMT. They tested the use of either a fully lexicalized or unlexicalized set of reordering rules on the task of MT between several language pairs including Arabic-to-English. The authors reported a very small improvement on the IWSLT 2004 and 2005 Arabic-to-English evaluation test sets. Habash [47] presented a reordering method for the Arabic-to-English MT task. He used a source Arabic dependency parse tree along with word alignment to automatically extract syntactic-level reordering rules from a parallel corpus. These rules have been used to reorder the Arabic training and testing data to match the target-side order. He investigated various alignment strategies and parsing representations and provided a comparative analysis of the different combinations of the investigated strategies. His approach gave a significant gain of 25% relative BLEU score when tested on an Arabic-to-English phrase-based MT baseline. Carpuat et al. [48, 49] used an Arabic noisy syntactic dependency parser to reorder verb-subject constructions into a pre-verbal subject-verb order. They tested their approach on an Arabic-to-English SMT baseline and reported a noticeable improvement in terms of both BLEU and TER scores. Bisazza et al. [50] tried to address the syntactic disfluencies that are found in the Arabic-to-English phrase-based SMT systems. They proposed a chunk-based reordering method that automatically reorders the Arabic verbs of the source-side sentences that follow a Verb–Subject–Object word order. Their method uses a feature-rich discriminative model to predict the likelihood of each possible verb reordering for a given Arabic source-side sentence. They reported a 1 BLEU point increase on the NIST-MT 2009 Arabic–English translation benchmark. Alqudsi et al. [57] proposed a method to handle word ordering problem in the context of Arabic-to-English MT. Their proposed method combines rule-based MT with the Expectation-Maximization (EM) algorithm. They used parallel data from the United Nations (Arabic–English) corpus. They trained their model using 632 sentence pairs and reserved 271 sentence pairs for testing. Their results showed an increase of up to 0.89 BLEU points over their RBMT baseline system.

English-to-Arabic. Elming and Habash [51] investigated the effectiveness of applying the reordering approach which was initially proposed to translate from English-to-Danish by Elming [52] on the task of English-to-Arabic MT. This approach attempts to learn probabilistic rules from a parallel corpus in a fully automatic way. They achieved an improvement in both manual and automatic evaluations. They also found that the rules that are learned from automatic alignments are more useful than those learned from the manual ones. Badr et al. [53] proposed a source-based syntactic reordering for the task of English-to-Arabic translation. Their reordering is carried out on the English source parse tree via handcrafted reordering rules that have been built based on their knowledge about the linguistic transformations that need to be accounted for when translating from English to Arabic. They reported some improvements in their phrase-based statistical MT baseline system. Hadj Ameur et al. [54] presented a POS-based preordering method to address both long- and short-distance reordering phenomena on the task of English-to-Arabic MT. They used syntactic unlexicalized reordering rules extracted automatically from a parallel corpus to reorder the English source-side language so that it matches that of the target Arabic one. They tested their method on the IWSLT2016 English-to-Arabic evaluation benchmark and reported more than 1 BLEU point increase in the overall translation results.

Between Arabic and Other Languages. Sadat and Mohamed [55] investigated several Arabic preprocessing schemes based on POS tagging and morphological segmentation on the task of Arabic-to-French MT. They used morphological rules to reduce the Arabic morphology to a level that makes it similar to the French one. They also used some reordering rules to match the source Arabic language with the target one on the syntactic level. Their tests on an Arabic–French SMT baseline showed that their morphological preprocessing was indeed useful. Their syntactic reordering rules, however, did not result in any significant improvement. Mohamed and Sadat [56] used handcrafted morphological reordering rules to reorder the source-side Arabic sentences in an Arabic-to-French translation task. Their rules attempt to reorder both the pronouns and verbs of the source-side Arabic sentences in a way that matches the target French language. They reported an improvement of around 1 BLEU point over their baseline statistical MT system.

5.1.3. Word Alignment

Some research studies attempted to improve the quality of word alignment as a means of improving the overall SMT quality.

Arabic-to-English. Ittycheriah and Roukos [58] presented a maximum entropy word alignment model for the task of Arabic-to-English MT. Their model was trained in a supervised way using annotated training data and compared to several other state-of-the-art word alignment algorithms such as IBM alignment model 1 [36, 107] and the HMM algorithm [108]. Even though a noticeable improvement has been obtained on the task of word alignment, it did not led to any significant gain in the overall MT results. Gao et al. [61] proposed a semi-supervised word alignment algorithm that combines a discriminative and a generative alignment model which they named EMDC (Expectation-Maximization Discrimination Constraint). The discriminative model finds high precision partial alignments and the generative one uses an Expectation-Maximization algorithm to impose additional constraints. Their tests on Chinese-to-English and Arabic-to-English translation tasks reported a consistent gain in the translation quality. Riesa and Marcu [62] presented a hierarchical search algorithm to address the problem of automatic word alignment. Their proposal introduces a forest of alignments from which they identify the best alignment points using a linear discriminative model that incorporates hundreds of features. Their test results on the task of Arabic-English word alignment showed a significant increase of 6.3 points in F-measure over a GIZA++ Model-4 baseline and a 1.1 BLEU score gain over a syntax-based SMT system. Hermjakob [60] proposed a method to improve the task of Arabic-English word alignment and translation by combining statistical and linguistic knowledge. Their method uses a bilingual lexicon produced by a statistical word aligner and a set of heuristic alignment rules generalized from a development corpus. Their proposal outperformed GIZA++ and LEAF aligners in terms of F-measure and produced a 1.3 BLEU score increase over a state-of-the-art syntax-based SMT baseline. Fossum et al. [59] presented a link deletion method to improve the word alignment produced by the GIZA++ toolkit. They used lexical, structural, and syntactic features to detect and delete weak alignment links produced by the GIZA++ aligner. Their tests on the tasks of Chinese-to-English and Arabic-to-English word alignment gave a significant F-measure increase, yet the gain in the overall translation results have not been as significant. Khemakhem et al. [63] proposed a method to improve the quality of Arabic-to-English SMT by including semantic-level knowledge. They first used a semantic word clustering method on the English side of the corpus to obtain semantic word classes. Then they linked these classes to Arabic using Arabic-English word alignment information. Their test results on the IWSLT 2008 and 2010 Arabic-English translation tasks showed up to a 1.4 BLEU score improvement over their statistical MT baseline.

English-to-Arabic. Ellouze et al. [64] proposed a hybrid approach to improve the alignment results of the GIZA toolkit. Their proposal uses linguistic features such as morpho-syntactic tags, syntactic patterns, and statistical features such as mutual information and harmonic mean. They trained their alignment model using an English-Arabic medical corpus and tested it on the Cambridge dictionary. They stated that their results showed an improvement in both the alignment and the translation quality. Berrichi and Mazroui [65] presented an alignment approach that uses morphosyntactic features (stem, lemma, and POS tags) for the task of English-to-Arabic MT. They evaluated their approach using a phrase-based SMT system as their baseline and reported a significant improvement in both the word alignment quality and the overall BLEU score results.

5.1.4. Language Models

A few research studies have focused on investigating the effectiveness of a language model that incorporates new morphological, syntactic, or semantic features.

Arabic-to-English. Brants et al. [66] investigated the impact of using a very large statistical English language model on the task of Arabic-to-English SMT. Their language model was trained on over 2 trillion tokens and yielding up to 300 billion n-grams. Their test results reported that the overall translation quality kept increasing gradually with the increase in the size of the language model even when reaching the largest size that they considered. Carter and Monz [67] used a large-scale discriminative language model to re-rank the n-best list translations generated by an SMT system. Their test results performed on NIST's Arabic-to-English MT-Eval benchmarks reported an improvement of 0.4 BLEU points over the state-of-the-art phrase-based SMT baseline that they considered. Niehues et al. [68] investigated the effect of using a bilingual language model that extends the translation model of a phrase-based SMT by including bilingual word context on the task of SMT. They tested their proposal on the tasks of German-to-English and Arabic-to-English MT and they reported an overall improvement of up to 1.7 BLEU points on the Arabic-to-English task.

English-to-Arabic. Khemakhem et al. [69] built an Arabic statistical feature-based language model that allows the incorporation of several grammatical features about each Arabic word. Their proposal was used to enhance the performance of an English-to-Arabic SMT system and they reported over 1-point BLEU score increase over the state-of-the-art phrase-based SMT baseline.

5.1.5. Other Arabic SMT Research Studies

Few researchers have investigated other morphological, syntactic, and semantic aspects to improve the quality of Arabic SMT systems.

Arabic-to-English. Habash [70] proposed several methods to address the problem of Out-of-Vocabulary (OOV) words in the task of Arabic-English SMT. His method used morphological, spelling, and dictionary enhancement; he also used a method for proper name transliteration. He reported a noticeable improvement over the state-of-the-art phrase-based SMT baseline in terms of BLEU score and a manual evaluation. Marton et al. [71] investigated the effect of adding soft constituent-level constraints to the Arabic source parse tree on the task of Arabic-to-English MT. They also used a feature weight optimization technique to handle the problem of selecting the best features. Their tests on an Arabic-to-English hierarchical phrase-based translation system showed substantial gains in performance.

English-to-Arabic. Toutanova et al. [72] used a specific inflection generation model to predict the correct inflections of a specific target language stems based on morphological and syntactic features extracted from both the source and target languages. Their proposed model was trained separately from the SMT baseline and tested on the task of translation of English to both Russian and Arabic. They reported an improvement of about 2 BLEU points on the task of English-to-Arabic translation.

5.2. Research Studies on Arabic Neural Machine Translation

The amount of research studies that have been devoted to the NMT paradigm has seen a very significant increase in the last few years. In this section, we classify the current research studies that have been accomplished in regards to Arabic NMT into three main categories:

1. Pre- and post-processing: attempt to improve the quality of NMT systems by using pre- or/and post-processing treatments.
2. Morphology, vocabulary, and factored NMT: investigates the incorporation of different linguistic knowledge sources into baseline NMT systems.
3. Multilingual and low-resource translation: attempt to use multilingual NMT under both rich- and low-resource settings.

5.2.1. Pre- and post-processing

Several research studies have been devoted to studying the effect of performing a pre- and/or post-processing treatments on Arabic NMT baselines.

Between Arabic and English. Sajjad et al. [73] investigated the effectiveness of three language-independent segmentations, namely: (1) Byte-Pair Encoding (BPE) [109], (2) Character-level Encoding [110], and (3) Character CNN [111]. They tested these segmentations on the Arabic-to-English and English-to-Arabic MT tasks and reported that the BPE segmentation produced the best results and even outperformed the state-of-the-art morphological segmentation (MADAMIRA[15]) on the

[15] `https://camel.abudhabi.nyu.edu/madamira/`

Arabic-to-English translation direction by a slim margin of 0.2 BLEU points. They also found that the character-level encoding methods perform drastically worse than both the BPE and the morphological segmentation ones lagging behind by more than two BLEU points on the English-to-Arabic evaluation tests that they performed. Oudah et al. [74] compared the effect of different segmentation schemes on neural and statistical Arabic-English MT models. Their results showed that the effect of the segmentation scheme is closely related to the type of the used translation model. They found that a morphology-based segmentation scheme such as the one used by the Arabic Treebank (ATB) has been beneficial to both the NMT and SMT models. However, the improvement was higher for the SMT models (reaching an increase of up to 3 BLEU points). They also found that the combination of the ATB with the BPE segmentation gave the best results for the SMT models but did not lead to an increase over the ATB segmentation for the NMT models. The overall conclusions were that ATB was the most useful segmentation for both SMT and NMT and that for SMT combining it with BPE can lead to an additional slight improvement. Hadj Ameur et al. [75] proposed a post-processing method for n-best list re-scoring in the context of English-to-Arabic and Arabic-to-English NMT. They used a set of features that cover the lexical, syntactic, and semantic aspects of the translation candidates (the n-best list candidates). The features that they used were extracted automatically from parallel corpora without needing any language-related tools. They also used the Quantum-behaved Particle Swarm Optimization (QPSO) algorithm[16] to optimize the weights of their incorporated features. Their system has been evaluated on multiple in-domain and out-domain Arabic-to-English and English-to-Arabic test sets, and they reported that their re-ranking results yield noticeable improvements of more than 1.5 BLEU points over their baseline NMT systems.

Between Arabic and Other Languages. Aqlan et al. [76] proposed a romanization system that converts Arabic scripts to subword units to deal with the unknown words problem on the task of MT between Arabic and Chinese. They investigated the effect of their approach on the NMT performance while using various segmentation scenarios. They performed extensive experiments on Arabic-to-Chinese and Chinese-to-Arabic translation tasks and showed that their proposed approach can effectively tackle the unknown words problem and improve the translation quality by up to 4 BLEU points.

[16] QPSO is a global-convergence-guaranteed optimization algorithm proposed by Sun et al. [112].

5.2.2. Morphology, Vocabulary, and Factored NMT

This section summarizes the research studies that have investigated the possibility of improving baseline NMT models via the incorporation of different linguistic knowledge sources.

Between Arabic and English. Ding et al. [77] tried to find the optimal vocabulary size for NMT models that uses subword units. They performed a wide range of experiments in which they varied the vocabulary size (the number of BPE merged operations) across several pairs of languages and reported the obtained results in terms of BLEU score. They found that for the Transformer-based Arabic-to-English and English-to-Arabic architectures the highest BLEU scores are obtained when the vocabulary size contains less than 1000 subword units. They reported a major drop in performance when the vocabulary size contains more than 8000 subword units. They also noted that the difference between the best and the worst performance is about 3 BLEU points. Ataman et al. [78] proposed a novel NMT decoding method that models word-formation via a hierarchical latent variable that simulates morphological inflection. Their method is aimed at morphologically rich and low-resource languages such as Arabic. Their proposal generates words one character at a time by combining two latent variables; the first is used to represent the lemmas and the second one for the inflectional features. They compared their proposal to subword and character-level decoding methods on the task of translation from English into three morphologically rich languages: Arabic, Czech, and Turkish. They reported a slight improvement of 0.51 BLEU points over the best performing baseline on the task English-to-Arabic translation. Ataman et al. [79] proposed a hierarchical decoding method for NMT that considers both words and characters when generating the translation. They compared their proposed method with different open-vocabulary subword-level techniques such as BPE across five pairs of languages with distinct morphological typologies. They showed that their hierarchical decoding model can give similar or even better results than the subword-level NMT models while using significantly fewer parameters. For the English-to-Arabic translation task, they reported an increase of up to 1.3 BLEU points over the baseline BPE subword-based NMT model. Liu et al. [80] proposed a novel method that allows the sharing of source and target word embedding features in the context of an NMT system. Their word embeddings are composed of two parts: shared features which are bilingual features used to improve the NMT's attention mechanism and private features that are used to capture the monolingual words features. The experiments that they have performed on five language pairs including Arabic-English showed significant performance increase over the Transformer baseline (an increase of up to 1.5 BLEU points for the Arabic-to-English translation direction) while using fewer model parameters. Shapiro and Duh [81] proposed a method that extends the word2vec word embeddings model by allowing it to include morphological lemmas from a language-specific morphological analyzer (MADAMIRA[17]). They showed that their proposed model outperformed word2vec on an Arabic word similarity task. They also performed experiments on Arabic-to-English translation tasks using the TED Talks data and found that the usage of morphological word embedding led to an improvement of 0.4 BLEU points over the original word2vec embeddings model and more than two BLEU points when compared to the random initialization.

Between Arabic and Other Languages. Belinkov et al. [82] performed several tests regarding NMT systems to identify the best possible morphological language-related representations. Their tests with different morphological processing on several languages such as French, German, Czech, Hebrew, and Arabic revealed that character-based representations tend to be better at learning morphology and that translating from a rich to a morphologically-poor language generally leads to better source-side representation. For instance, the character-based segmentation showed an improvement of more than 3 BLEU points over the word-based one on the Arabic-to-English translation direction, while on the English-to-Arabic their results were very similar. Ataman and Federico [83] investigated the use of vocabulary reduction techniques to improve the quality of neural machine translation when dealing with morphologically-rich languages. They tested two unsupervised vocabulary reduction methods: Byte-pair encoding and Linguistically-Motivated Vocabulary Reduction (LMVR) [113]. They compared the two methods on ten translation directions that involve English and five morphologically-rich languages: Arabic, Czech, German, Italian, and Turkish. They showed that the performance of the subword segmentation method was better for the majority of the tested language pairs. As to the Arabic language, they found that the LMVR segmentation was better than the BPE for both Arabic-to-English and English-to-Arabic by around one BLEU point. García-Martínez et al. [84] investigated the effect of using linguistic factors on the target-side of an Arabic-to-French factored NMT model [18]. Two pieces of information were predicted by their FNMT model at decoding time, the lemma, and the concatenation of the following factors: POS tag, tense, gender, number, person, and the case information. Their training was done using a small or large parallel training dataset to simulate low-resource and rich-resource behaviors, respectively. They also investigated the usage of BPE segmentation for both their Factored and standard NMT architectures. Their evaluation results on several test sets showed that the factored NMT models were far better under low-resource conditions by an improvement of around 3 to 6 BLEU points over the baseline NMT. They also found that combining factors with subword BPE units achieved the best performance when trained under their rich-resource settings.

5.2.3. Multilingual and Low-resource Translation

Some research studies were interested in the usage of multilingual NMT under both rich- and low-resource settings.

Between Arabic and English. Nishimura et al. [85] examined the usefulness of multi-source NMT which incorporates multiple source inputs (from different languages). They also presented a method to use incomplete multilingual parallel corpora in which some source or target translations can be missing. They used UN6WAY multilingual corpus from which they selected Spanish, French, and Arabic as the source languages and English as the target language. Their proposed multi-source NMT model achieved an increase of up to 5.6 BLEU points over the best one-to-one NMT baseline. Tan et al. [86] proposed a framework in which several languages are grouped into different clusters, each one trained as a multilingual model. They tested two approaches for language clustering: (1) using human-knowledge, thus clustering languages according to their families; and (2) using language embeddings, thus, representing each language via an embedding vector and then clustering them according to those embeddings. They tested their two clustering methods on the task of MT from 23 languages (including Arabic) to English.

[17] `https://camel.abudhabi.nyu.edu/madamira/`

[18] Factored NMT architectures consider several linguistic factors (e.g. part-of-speech, case, number, gender) of the source and/or the target language to improve the overall NMT quality [114, 115].

Their first clustering method placed Arabic and Hebrew in the same "Afroasiatic" cluster and the second one placed Arabic, Persian, and Hebrew in the same cluster. Their test results showed that the second clustering method was better almost in all scenarios, leading to an improvement of 0.25 BLEU points in the case of Arabic-to-English. Liu et al. [87] presented mBART (Multilingual BART, which is an extension of the original BART[19]) a denoising auto-encoder that they pretrained on several monolingual language corpora. They have shown that using mBART has led to an increase in performance of up to 12 BLEU points for some low resource sentence-level translation tasks and an increase of around 5 BLEU points for several document-level translation tasks. As far as Arabic is concerned, their mBART25 (pretrained on 25 languages) has led to an increase of 10.1 BLEU points on the task of Arabic-English MT. Aharoni et al. [88] presented a large multilingual NMT translating 102 languages to and from English. Their test results have been reported on the TED talks multilingual test set, and they showed that their system has been particularly effective under low resource settings. As far as Arabic is concerned, their multilingual NMT achieved a BLEU score increase of 2 to 3 points over the one-to-one NMT models on both English-to-Arabic and Arabic-to-English translation directions.

Between Arabic and Other Languages. Almansor and Al-Ani [89] presented a character-based hybrid NMT model that combines both recurrent and convolutional neural networks. They trained their model on a very small portion of the TED parallel corpora containing only 90K sentence pairs. They tested their model on the IWSLT 2016 Arabic-to-English and English-to-Vietnamese evaluation sets and they reported noticeable improvements in comparison to a standard word-based NMT model. For the case of English-to-Arabic translation, the improvement in BLEU score exceeded 10 BLEU points while the word-based NMT model completely failed to train using their very small parallel training corpus. Ji et al. [90] proposed a transfer learning approach to handle the translation of low-resource languages based on cross-lingual pretraining. Their method trains a universal encoder on several source languages using a shared feature space. Once the universal encoder is pretrained on the monolingual source languages data, the whole NMT model will then be trained using parallel data and used in zero-shot translation scenarios. Their tests on Europarl (involving French, English, Spanish, German, and Romanian languages) and MultiUN (involving Arabic, Spanish, and Russian) test sets showed that their approach significantly outperforms both pivot-based and multilingual NMT baselines. As far as Arabic is concerned, their experiments on the MultiUN test set for the tasks of Spanish and Russian translation from and to Arabic reported an improvement that ranges from 1 to 3 BLEU points over their multilingual NMT baseline model.

[19]BART [116] is a denoising sequence-to-sequence autoencoder trained to denoise and reconstruct texts that have been corrupted by using arbitrary noising functions.

5.2.4. Comparing Neural and Statistical MT Performance

Some research studies attempted to compare the performance of neural and statistical Arabic MT systems.

Between Arabic and English. Almahairi et al. [91] developed an NMT system for the task of Arabic-to-English translation and compared its performance to that of a phrase-based SMT system. They performed extensive tests using several Arabic preprocessing configurations and found that the phrase-based and neural translation systems give similar results and that a proper Arabic preprocessing has a positive impact on both of them. They also observed that their NMT system performs significantly better when evaluated on out-of-domain test sets. Junczys-Dowmunt et al. [92] performed a large comparison between phrase-based and neural MT systems for several language pairs including the Arabic-to-English and English-to-Arabic translation directions. They reported that the NMT results were on par or better than those of the phrase-based SMT. For some languages, a very slight increase has been observed; however, when the Arabic language was involved, noticeable increases that range between 1 and 9 BLEU points were reported over the phrase-based SMT systems. For both Arabic-to-English and English-to-Arabic a 3-point increase in BLEU score has been observed.

Between Arabic and Other Languages. Belinkov and Glass [93] compared the performance of phrase-based and neural MT systems on the task of Arabic-to-Hebrew translation. They tested the effect of tokenization on both NMT and PSMT systems and they also tested the impact of character-level models on the NMT system. The preprocessing step resulted in a significant gain for both NMT and PSMT. They reported that their NMT gave better results when compared to phrase-based MT and that char-based models led to a one BLEU point improvement over their baseline NMT system.

5.3. Research Studies on Arabic Rule-based Machine Translation

Very limited research studies were developed using the classical rule-based methods to address the task of Arabic MT.

Arabic-to-English. Salem et al. [94] investigated the process of developing a rule-based lexical framework specialized in Arabic language translation by using Role and Reference Grammar (RRG) models. They described the difficulty of incorporating the Arabic language characteristics in such a framework when developing an Arabic-to-English translation system. Shirko et al. [95] developed an MT system that translates Arabic noun phrases into English using a transfer-based approach. They tested their system by translating 88 thesis and journal paper titles from the computer science domain and reported an accuracy of 94.6%.

English-to-Arabic. Soudi et al. [96] described a research project aiming to build an English-to-Arabic interlingual MT system. They proposed a mapping procedure that allows the mapping of different semantic concepts from English to Arabic in the interlingual representation. They also addressed some of the differences between English and Arabic such as agreement in number. They provided a simple example illustrating their proposal. Shaalan et al. [97] proposed a method to construct a transfer-based English-to-Arabic MT system specialized in translating complex English noun phrases into Arabic. Their system follows the analysis, transfer, and generation steps to transform the English noun phrases into Arabic. Their proposal was tested on the task of translating theses titles taken from the computer science domain, and they reported a 92% translation accuracy on a holdout test set. Al Dam and Guessoum [98] investigated the effectiveness of using an artificial neural network (ANN) in a transfer-based English-to-Arabic MT system. They used a feed-forward neural network with two hidden layers as a transfer module which learns to transfer a tagged English sentence into a tagged Arabic one. Their tests showed that 56% of the test sentences were perfectly transferred and that 64.5% of them had at least 60% correct tags.

5.4. Research Studies on the Evaluation of Arabic MT Systems

Some researchers proposed new methods to better evaluate the quality of Arabic MT systems.

Arabic-to-English. Hadla et al. [99] evaluated two Arabic-to-English MT systems namely Google Translate and Babylon. They used a corpus of more than 1000 Arabic-English sentence pairs which associate two reference English translations for each source Arabic sentence. Their Arabic sentences were distributed among four sentence functions: declarative, interrogative, exclamatory, and imperative. Their experimental results reported in terms of the BLEU-score showed that Google's translation quality was better than Babylon's.

English-to-Arabic. Guessoum and Zantout [100] proposed a methodology for performing a semi-automatic evaluation of lexicons in the context of MT. Their method takes into consideration the importance of a given word in each possible domain; thus they gave an importance weight to each word in the lexicon based on its morphological properties and its specific sense. They used their proposed methodology to test the lexicons of three English-to-Arabic MT systems: Al-Mutarjim Al-Arabey, Arabtrans, and Al-Wafy[20]. They reported that all the tested systems gave a lexical coverage of more than 93% which they described as "acceptably good". In their later work, Guessoum and Zantout [101] proposed a generalization to their lexicon-based evaluation in which the central idea of word sense weights was generalized to account for grammatical and semantic correctness. Their methodology was tested on four English-to-Arabic MT systems ATA, Arabtrans, Ajeeb, and Al-Nakel[21]. Their test results showed poor performance for all the evaluated systems that vary between 32% and 64% correctness on grammatical coverage, and between 51% and 84% on semantic correctness. Adly and Al Ansary [102] proposed an Interlingua-based approach for the evaluation of English-to-Arabic MT systems. They used the Universal Networking Language (UNL), a formal declarative language that represents textual data in a semantic way. They compared their evaluation method with some commonly used automatic evaluation metrics such as BLEU, F-1, and F-mean and reported that their proposal was better especially when dealing with sentences that have a complex structure. Al-Rukban and Saudagar [103] compared the performance of some English-to-Arabic translation systems using two metrics: BLEU and General Text Matcher (GTM). They considered three systems:

[20] Al-Wafi and Al-Mutarjim Al-Arabey are commercial Arabic MT systems developed by ATA Software Technology Inc, and Arabtrans is a commercial Arabic MT system developed by ArabNet.

[21] Al-Nakel is a commercial Arabic MT system developed by CIMOS.

Google Translator, Bing Translator, and Golden Alwafi. Their tests showed that Golden Alwafi was better in terms of quality as measured using BLEU, but that Google Translator surpassed it when using the GTM metric. El Marouani et al. [105] investigated the impact of using linguistic features to evaluate the Arabic output of English-to-Arabic translation systems. Their proposed features consider semantic aspects from both the source and the target languages. The tests performed on a medium-sized corpus showed that their features helped improve the correlation between automatic and manual assessments. Guzmán et al. [104] investigated the effect of incorporating some embedding features that have been obtained from different lexical and morpho-syntactic linguistic representations on the task of MT evaluation. They used a pairwise feed-forward neural network that takes the linguistically motivated embeddings of two possible translations and picks the best one among them. Their tests on the task of MT from English to Arabic showed that their proposal outperforms state-of-the-art evaluation metrics by an increase of over 75% in the correlation with human judgment as reported on a pairwise MT evaluation quality task.

6. Arabic MT Resources and Tools

The availability of linguistic corpora and tools has a huge impact on the overall MT quality. In this section, we will try to highlight the most substantial tools and resources that are available in the field of Arabic MT[22].

6.1. Arabic MT Parallel Training Datasets

Currently, a large amount of parallel corpora exists for the task of translation between Arabic and most of the top spoken languages in the world such as English, Chinese, French, Spanish, etc. The largest collections of freely available parallel corpora can be obtained from the OPUS website[23]. In the following, we will cite the two largest OPUS parallel corpora that include the Arabic language. We encourage the readers to visit the OPUS website to check their full collection of parallel Arabic corpora.

1. The United Nations Parallel Corpus v1.0 [117][24]: a freely available parallel corpus built from manually translated UN documents between the years 1990 and 2014 for the six official UN languages: Arabic, Chinese, English, French, Russian, and Spanish. The corpus contains around 20 million sentence pairs for each translation direction between the concerned UN languages.
2. OpenSubtitles [118][25]: a large corpus built from movie and TV subtitles gathered from 60 languages including Arabic. The corpus contains more than 30 million English-Arabic sentence pairs and around 20 million sentence pairs for the tasks of translation between Arabic and most of the major spoken-languages such as Chinese, French, Russian, Spanish, etc.

Another major source that offers large collections of corpora is the Linguistic Data Consortium (LDC)[26]. LDC offers large translation data sets of Arabic newswire, weblog, forums, broadcast transcripts, and newsgroup texts. The majority of their data sets are translated manually validated rigorously by translation experts. In the following, we will cite some important LDC parallel corpora that include the Arabic language, but we note that there are so many, thus we encourage the readers to visit their website.

1. Arabic Broadcast News Parallel Text (LDC2007T24, LDC2008T09, and LDC2015T07): contains English translations of several hours of Arabic broadcast news (more than 100k Arabic words).
2. Arabic Newsgroup Parallel Text (LDC2009T03 and LDC2009T09): contains English translation of texts (more than 300k Arabic words) driven from forums posts, discussion groups, etc.
3. Arabic Newswire English Translation Collection (LDC2009T22): contains texts from Arabic Newswire: Agence France Presse (France), An Nahar (Lebanon), and Assabah (Tunisia), along with their English translations (more than 500k Arabic words).
4. TRAD Arabic-French Parallel Text (LDC2018T13, LDC2018T21): contains French translations of a subset of more than 30k Arabic words developed under the PEA-Trad project[27].

6.2. Arabic MT Evaluation Datasets

Both free and paid test sets are available for the evaluation of Arabic MT systems. The free test sets generally

[22] We note that all the details and statistics regarding the mentioned data sets can be found in their provided links.
[23] `http://opus.nlpl.eu/`
[24] `http://opus.nlpl.eu/UNPC-v1.0.php`
[25] `http://opus.nlpl.eu/OpenSubtitles-v2018.php`
[26] `https://www.ldc.upenn.edu/`
[27] `http://www.elra.info/en/projects/archived-projects/pea-trad/`

provide only one reference translation for each source sentence, while the paid test sets provide multiple reference translations for each source sentence which allows a more reliable and robust evaluation. In the following, we will cite the most used and freely available Arabic MT evaluation test sets.

1. Arab-Acquis Dataset [119][28]: is an evaluation set created from the European Union's Acquis Communautaire corpus which involves law-related textual data. Arab-Acquis dataset allows the evaluation of MT systems that translate between Arabic and 22 European languages.
2. International Workshop on Spoken Language Translation [120][29]: also known as IWSLT, is a yearly scientific workshop that offers an evaluation campaign for spoken language translation. They released several test sets for the evaluation of Arabic-to-English and English-to-Arabic MT systems which are IWSLT 2012, IWSLT 2013, IWSLT 2014, IWSLT 2015, IWSLT 2016, and IWSLT 2017.
3. The United Nations Parallel Corpus v1.0 Test Set [117][30]: is an evaluation data set driven from the United Nations Parallel Corpus that allows the evaluation of translation between the six official UN languages: Arabic, Chinese, English, French, Russian, and Spanish.

Besides the aforementioned free test sets, many paid test sets are available for Arabic MT. In the following, we will cite the most important ones among them.

1. NIST Open Machine Translation Evaluation: offers a large collection of test sets (that can be obtained from the LDC website) intended for the evaluation of MT systems quality across several languages such as Arabic, English, Korean, Farsi, Urdu, Chinese, etc. The NIST MT evaluation test sets that involve the Arabic language are the following:
    - For the tasks of Chinese-to-English and Arabic-to-English translation: NIST OpenMT 2002 (LDC2010T10), NIST OpenMT 2003 (LDC2010T11), NIST OpenMT 2004 (LDC2010T12), NIST OpenMT 2005 (LDC2010T14), NIST OpenMT 2006 (LDC2010T17), NIST OpenMT 2008 (LDC2010T21).
    - For the tasks of Arabic-to-English and Urdu-to-English translation: NIST OpenMT 2009 (LDC2010T23).
    - For the tasks of Arabic-to-English, Chinese-to-English, Dari-to-English, Farsi-to-English, and Korean-to-English translation: NIST OpenMT 2012 (LDC2013T03).
2. TRAD Arabic-French Newspaper Parallel Test set[31]: an Arabic-French evaluation test set created from "Le Monde Diplomatique"[32] articles of the year 2012. It contains two parts: TRAD Arabic-French Newspaper Parallel corpus Test set 1 (ELRA-W0098) and TRAD Arabic-French Newspaper Parallel corpus Test set 2 (ELRA-W0100).

6.3. Monolingual Arabic Datasets

Monolingual corpora are very helpful for the task of MT and the field of NLP in general. They can be used to train word and sentence embedding models, language models, subword segmentation models, etc. Due to the importance of the Arabic language (one of the five most spoken languages in the world[33]), a good number of large-sized monolingual Arabic text corpora are currently available. In the following, we will try to mention the most substantial ones among them.

- Arabic Gigaword (LDC2003T12, LDC2006T02, LDC2007T40, and LDC2009T30): the four parts of the Arabic Gigaword contain a total of around two trillion Arabic tokens (more than 20GB of raw Arabic texts) gathered from newswire agencies such as Agence France Presse, Al Hayat News Agency, and Al Nahar News Agency.

- Arabic Newswire Part 1 (LDC2001T55): contains Arabic articles (a total of 76 million tokens) gathered from the French Press Agency.

Besides the LDC paid corpora, a large number of freely monolingual data sets are also available.

- Tashkeela [121][34]: a corpus containing 75.6 million vocalized Arabic words.

- KSUCCA Corpus [122][35]: a collection of classical Arabic texts concerning the period from the pre-Islamic era to the fourth Hijri century.

[28] https://camel.abudhabi.nyu.edu/arabacquis/
[29] https://wit3.fbk.eu/
[30] https://conferences.unite.un.org/UNCorpus/

[31] http://catalog.elra.info/
[32] https://www.monde-diplomatique.fr/
[33] https://www.conservapedia.com/List_of_languages_by_number_of_speakers
[34] https://sourceforge.net/projects/tashkeela/
[35] https://sourceforge.net/projects/ksucca-corpus/

- The International Corpus of Arabic [123]: 100 million words Arabic corpus extracted from different sources such as newspapers and web articles; it covers different domains such as science, literature, and politics.

- A collection of Arabic Corpora[36]: a website providing free access to multiple Arabic monolingual corpora such as "Ajdir Corpora" (113 million words), "Open Source Arabic Corpora" (20 million words), "Watan corpus" (12 million words) and "Khaleej corpus" (3 million words) [124, 125].

We note that besides the mentioned monolingual corpora, it is also possible to build web crawlers to extract Arabic texts automatically from Arabic news websites, Wikipedia, and other online sources. Also, all the aforementioned parallel Arabic corpora can be used as monolingual ones by just considering the sentences of their Arabic-side.

6.4. Arabic Treebanks

Treebanks are fully parsed corpora that are linguistically annotated at both the sentence and word levels. Treebanks are very crucial for many NLP tasks such as Part-of-Speech tagging, Named Entity Recognition, Dependency and Constituency Parsing, Relation Extraction, Machine Translation, and Question Answering [24]. In the following (Table 7), we will try to cite the most substantial Arabic Treebanks that are currently available.

Table 7: Arabic Treebanks

Name	Ownership	Availability	Number of tokens
Arabic Treebank Part 1 v 4.1 (LDC2010T13)	LDC	Paid	145,386
Arabic Treebank: Part 2 v 3.1 (LDC2011T09)	LDC	Paid	144,199
Arabic Treebank: Part 3 v 3.2 (LDC2010T08)	LDC	Paid	339,710
Arabic Treebank: Part 4 v 1.0 (LDC2005T30)	LDC	Paid	1,000,000
Prague Arabic Dependency Treebank 1.0 (LDC2004T23)	LDC	Paid	113,500
OntoNotes Release 5.0 (LDC2013T19)	LDC	Free	300,000
The Quranic Arabic Corpus	Open Source	Free	77,430
Nemlar Corpus	Nemlar Project	Free	500,000
Universal Dependencies Treebank	Open Source	Free	738,889

[36]`http://aracorpus.e3rab.com/index.php?content=english`

The LDC Treebanks listed in Table 7 are all available on the LDC website[37]. The Universal Dependencies Treebank can be obtained from the Universal Dependencies website[38]. The Nemlar Corpus [126, 127] can be obtained from the website of the Oujda NLP research group[39]. We note that at the time of writing this survey the annotation of the Quranic Arabic Corpus[40] was not completed. A total of 30,895 out of the 77,430 words (around 40%) have been annotated, and the remaining part is still undergoing the annotation process.

6.5. Arabic MT Tools

In this section, we will present the most relevant tools that are important for the training and evaluation of both statistical and neural MT systems.

Table 8 summarizes some of the most substantial MT tools that can help train and evaluate MT systems. However, given that the field of NLP is currently seeing an unprecedented rise in the number of available tools and utilities that keep appearing each day, we encourage the reader to always visit GitHub and check for new tools and updates.

7. Discussion

In Section 5, we summarized the most relevant research studies that have been developed in the field of Arabic MT (see Table 6). These studies have been categorized according to the translation approach they used along with the MT problem/sub-problem that they tackled. In this section, we will highlight the most important research directions that have been investigated in the studies that have been made in regards to Arabic MT and give our remarks, observations, and comments about them.

After a careful analysis of what has been done in the field of Arabic MT from its early days up until now, we can make the following observations:

- The Arabic MT studies that have been accomplished so far have primarily been devoted to translating Arabic to English; English to Arabic translation has been of secondary importance. As to other languages than English, there are only very

[37]`https://www.ldc.upenn.edu/`

[38]`https://universaldependencies.org/treebanks/ar_nyuad/index.html`

[39]`http://oujda-nlp-team.net/en/corpora/nemlar-corpus-en/`

[40]`http://corpus.quran.com/`

Table 8: Some relevant tools for training statistical and neural MT systems

Task	Tool's Name	Developed using	Main Contributors	Description
End-to-end NMT	OpenNMT[41]	Python/Pytorch and TensorFlow	Harvard, Systran, Ubiqus	OpenNMT [128] is an open-source framework for sequence learning and neural machine translation. It offers a very wide variety of models and options for training and testing different NMT models.
	Fairseq[42]	Python/Pytorch	Facebook Research	Fairseq [129] is a sequence modeling toolkit that implements different models for translation, language modeling, summarization, and text generation.
	Tensor2Tensor[43]	Python / TensorFlow	Google	Tensor2Tensor [130] is a library that implements deep learning models for MT and several other NLP tasks.
End-to-end SMT	Moses[44]	C++ and Perl	Edinburgh & Charles Universities	Moses [131] is a framework for training and testing SMT systems.
	Phrasal[45]	Java	Stanford University	Phrasal [132] is a framework that aims for the fast training of both traditional MT models and large-scale discriminative translation models.
Word Segmentation	Subword-nmt[46]	Python	Samsung Electronics	Subword-nmt is a GitHub repository that implements a set of preprocessing scripts for the training of subword unit (BPE) [133] segmentation models.
	SentencePiece[47]	Python	Google	SentencePiece [133] is an unsupervised text segmentation and desegmentation tool that implements several subword units algorithms.
Word Alignment	GIZA++[48]	C++	Johns-Hopkins University	GIZA++ [134] is a tool for learning statistical word-to-word alignment models from parallel corpora.
	FastText Multilingual[49]	Python	Babylon Health UK	FastText Multilingual [135] is a tool that can be used to align two language vocabularies form their respective monolingual pretrained embeddings.
Word Embedding	Gensim[50]	Python	RARE Technologies Ltd.	Gensim [136] is a library that offers a parallel implementation of many algorithms that can be used to train word embedding models.
	FastText[51]	Python	Facebook	FastText [137] is a GitHub repository that provides pretrained word embeddings for many languages including Arabic.
MT Evaluation	NLG Evaluation[52]	Python	Microsoft	NLG Evaluation [138] is a framework that implements several MT evaluation metrics such as BLEU, METEOR, and ROUGE.

few studies (for some languages no research studies can be found at all) even though parallel corpora are currently available for the task of MT between Arabic and many other languages.

- When MT was mainly rule-based, work on Arabic was very bare; by the time interest grew for Arabic MT, MT had largely moved to data-driven translation methods. Currently, the MT research studies are focused heavily on the data-based approaches (mainly SMT and NMT) that do not require any linguistic expertise, thus, rule-based MT methods which are extremely demanding in both cost and human effort [97, 94, 95] keep getting less and less attention over time.

- Arabic morphological analysis has been one of the most studied aspects in the field of Arabic MT. Indeed, the main concern when dealing with the Arabic language is its complex and rich morphology which is substantially different from that of Indo-European languages (such as English). These studies showed that Arabic tokenization and morphological segmentation can lead to some significant improvements in the overall translation results [8, 41, 43].

- Syntactic word reordering is also another very heavily studied aspect in the context of Arabic SMT. Indeed, the Arabic language is known to have a free word order which permits several possible word orderings, a thing that is not always permitted in other languages that tend to require a specific ordering. The majority of these reordering methods reported some clear gains in the overall translation results [47, 50, 54].

- Word alignment is also a subtask of SMT; it has been investigated by the Arabic MT research community. The proposed approaches attempted to improve it by introducing new linguistic features to boost the alignment quality. Even though some studies have managed to significantly improve the alignment results the impact of this gain on the

[41] https://opennmt.net/
[42] https://github.com/pytorch/fairseq
[43] https://github.com/tensorflow/tensor2tensor
[44] http://www.statmt.org/moses/
[45] https://github.com/stanfordnlp/phrasal
[46] https://github.com/rsennrich/subword-nmt
[47] https://github.com/google/sentencepiece
[48] https://github.com/moses-smt/giza-pp
[49] https://github.com/Babylonpartners/fastText_multilingual
[50] https://radimrehurek.com/gensim/
[51] https://github.com/facebookresearch/fastText/blob/master/docs/crawl-vectors.md
[52] https://github.com/Maluuba/nlg-eval

overall translation results has not always been as significant [58, 59].

- Feature-rich language models also have been investigated in the context of Arabic MT, yet the few studies that have adopted this approach have used different features and integration methodologies. Thus, drawing a clear conclusion about the effectiveness of this path is not trivial since the impact of these models is completely dependent on the considered features.

- NMT is the newest emerging paradigm for MT; yet, a considerable amount of research studies have been recently made with regard to it for the Arabic language. These studies have reported very encouraging results which were often better than the SMT-based ones, especially when tested on out-of-domain data [91, 92].

- Multilingual MT approaches that use a single model to translate between multiple languages have shown very promising translation results. Indeed, recent studies demonstrated their effectiveness not only for low-resource languages but also for languages that have large parallel data such as Arabic. The studies also showed that these models were capable of capturing shared representational features across languages, thus offering better transfer capabilities which lead to larger gains in translation quality [85, 88, 86].

8. Conclusion

In this survey, we have provided a comprehensive overview of the different research studies that have been proposed in the field of Arabic MT. To summarize, we first introduced the Arabic language, its characteristics along with some of its translation difficulties. Then, we presented the MT paradigms and highlighted some of their strengths and weaknesses. Next, we provided an overview of the important research studies that have been made so far on Arabic MT in a categorized way that allows the reader to distinguish with ease the different research axes and contributions that have been proposed. Finally, we provided a quick summary of the most relevant studies in each distinct area and discussed their effectiveness and shortcomings. Arabic MT is still not at the human level, yet it is getting better and better over time. In addition to the advances that have been achieved using SMT, NMT appears to be even a more promising path that can lead to further improvements.

Having done this survey, we can now highlight some promising research directions in the field of MT in general and Arabic MT in particular:

- Finding new ways of merging the advantages of neural and statistical machine translation in a single framework.

- Training NMT models with minimal or no data using zero-shot and transfer learning methods has shown very promising results for the tasks of MT across several pairs of languages. We believe that these kinds of methods need to be investigated thoughtfully as they can be very helpful for Arabic MT.

- Exploring the document-level neural translation methods which can take into account linguistic phenomena that go beyond the sentence-level boundaries, such as pronoun resolution, can solve many problems for Arabic MT.

- The incorporation of linguistic features to the NMT framework (such as factored and linguistically motivated models) have shown very promising results for the task of MT between several language pairs; thus, an in-depth exploration of these methods for Arabic MT may lead to substantial improvements.

- The usage of multilingual MT methods can lead to noticeable improvements for the task of Arabic MT as they are capable of capturing shared linguistic features across languages, thus offering better transfer capabilities.

- Many Arabic-related linguistic problems still need much investigation as they are still causing considerable difficulties for current Arabic MT systems. The main ones are the translation of Arabic named entities, idiomatic expressions, ambiguous words, out-of-vocabulary words, and Arabic poetry.

References

[1] W. J. Hutchins, Machine translation: A brief history, in: Concise history of the language sciences, Elsevier, 1995, pp. 431–445.

[2] W. J. Hutchins, Machine translation: past, present, future, Ellis Horwood Chichester, 1986.

[3] W. J. Hutchins, H. L. Somers, An introduction to machine translation, volume 362, Academic Press London, 1992.

[4] P. Koehn, Statistical machine translation, Cambridge University Press, 2009.

[5] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv preprint arXiv:1409.0473 (2014).

[6] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., Google's neural machine translation system: Bridging the gap between human and machine translation, arXiv preprint arXiv:1609.08144 (2016).

[7] M. Alkhatib, K. Shaalan, The key challenges for Arabic machine translation, in: Intelligent Natural Language Processing: Trends and Applications, Springer, 2018, pp. 139–156.

[8] N. Habash, F. Sadat, Arabic preprocessing schemes for statistical machine translation, in: Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers, Association for Computational Linguistics, 2006, pp. 49–52.

[9] A. Alqudsi, N. Omar, K. Shaker, Arabic machine translation: a survey, Artificial Intelligence Review 42 (2014) 549–572.

[10] N. Y. Habash, Introduction to Arabic natural language processing, Synthesis Lectures on Human Language Technologies 3 (2010) 1–187.

[11] K. C. Ryding, Arabic: A linguistic introduction, Cambridge University Press, 2014.

[12] A. Farghaly, K. Shaalan, Arabic natural language processing: Challenges and solutions, ACM Transactions on Asian Language Information Processing 8 (2009). URL: `https://doi.org/10.1145/1644879.1644881`. doi:doi:10.1145/1644879.1644881.

[13] C. S.-H. Ho, P. Bryant, Learning to read Chinese beyond the logographic phase, Reading research quarterly 32 (1997) 276–289.

[14] L. H. Tan, J. A. Spinks, G. F. Eden, C. A. Perfetti, W. T. Siok, Reading depends on writing, in Chinese, Proceedings of the National Academy of Sciences 102 (2005) 8781–8785.

[15] M. S. Hadj Ameur, Y. Moulahoum, A. Guessoum, Restoration of arabic diacritics using a multilevel statistical model, in: A. Amine, L. Bellatreche, Z. Elberrichi, E. J. Neuhold, R. Wrembel (Eds.), Computer Science and Its Applications, Springer International Publishing, Cham, 2015, pp. 181–192.

[16] E. Abuelyaman, L. Rahmatallah, W. Mukhtar, M. Elagabani, Machine translation of Arabic language: challenges and keys, in: Intelligent Systems, Modelling and Simulation (ISMS), 2014 5th International Conference on, IEEE, 2014, pp. 111–116.

[17] A. Farghaly, K. Shaalan, Arabic natural language processing: Challenges and solutions, ACM Transactions on Asian Language Information Processing (TALIP) 8 (2009) 14.

[18] N. Habash, O. Rambow, Arabic diacritization through full morphological tagging, in: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers, Association for Computational Linguistics, 2007, pp. 53–56.

[19] M. Diab, M. Ghoneim, N. Habash, Arabic diacritization in the context of statistical machine translation, in: Proceedings of MT-Summit, 2007.

[20] C. J. Fillmore, B. T. Atkins, Describing polysemy: The case of 'crawl', Polysemy: Theoretical and computational approaches (2000) 91–110.

[21] V. Kordoni, I. Simova, Multiword expressions in machine translation., in: LREC, 2014, pp. 1208–1211.

[22] I. A. Sag, T. Baldwin, F. Bond, A. Copestake, D. Flickinger, Multiword expressions: A pain in the neck for nlp, in: International Conference on Intelligent Text Processing and Computational Linguistics, Springer, 2002, pp. 1–15.

[23] M. Constant, G. Eryiğit, J. Monti, L. Van Der Plas, C. Ramisch, M. Rosner, A. Todirascu, Multiword expression processing: A survey, Computational Linguistics 43 (2017) 837–892.

[24] D. Jurafsky, J. H. Martin, Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition, 2nd Edition, Prentice Hall series in artificial intelligence, Prentice Hall, Pearson Education International, 2009. URL: `https://www.worldcat.org/oclc/315913020`.

[25] K. Shaalan, A survey of Arabic named entity recognition and classification, Computational Linguistics 40 (2014) 469–510.

[26] M. S. H. Ameur, F. Meziane, A. Guessoum, Arabic machine transliteration using an attention-based encoder-decoder model, Procedia Computer Science 117 (2017) 287 – 297. URL: `http://www.sciencedirect.com/science/article/pii/S1877050917321774`. doi:doi:https://doi.org/10.1016/j.procs.2017.10.120, arabic Computational Linguistics, Dubai.

[27] M. R. Costa-Jussa, J. A. Fonollosa, Latest trends in hybrid machine translation and its applications, Computer Speech & Language 32 (2015) 3–10.

[28] M. Nagao, A framework of a mechanical translation between japanese and English by analogy principle, Artificial and human intelligence (1984) 351–354.

[29] P. Brown, J. Cocke, S. D. Pietra, V. D. Pietra, F. Jelinek, R. Mercer, P. Roossin, A statistical approach to language translation, in: Proceedings of the 12th conference on Computational linguistics-Volume 1, Association for Computational Linguistics, 1988, pp. 71–76.

[30] P. F. Brown, J. Cocke, S. A. D. Pietra, V. J. D. Pietra, F. Jelinek, J. D. Lafferty, R. L. Mercer, P. S. Roossin, A statistical approach to machine translation, Computational linguistics 16 (1990) 79–85.

[31] P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, R. L. Mercer, The mathematics of statistical machine translation: Parameter estimation, Computational linguistics 19 (1993) 263–311.

[32] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078 (2014).

[33] M. Schuster, K. K. Paliwal, Bidirectional recurrent neural networks, IEEE Transactions on Signal Processing 45 (1997) 2673–2681.

[34] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.

[35] K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio, On the properties of neural machine translation: Encoder-decoder approaches, arXiv preprint arXiv:1409.1259 (2014).

[36] F. J. Och, H. Ney, Improved statistical alignment models, in: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2000, pp. 440–447.

[37] N. Habash, B. Dorr, C. Monz, Symbolic-to-statistical hybridization: extending generation-heavy machine translation,

Machine Translation 23 (2009) 23–63.
[38] F. Xia, M. McCord, Improving a statistical mt system with automatically learned rewrite patterns, in: Proceedings of the 20th international conference on Computational Linguistics, Association for Computational Linguistics, 2004, p. 508.
[39] D. Groves, A. Way, Hybrid data-driven models of machine translation, Machine Translation 19 (2005) 301–323.
[40] Y.-S. Lee, Morphological analysis for statistical machine translation, in: Proceedings of HLT-NAACL 2004: Short Papers, Association for Computational Linguistics, 2004, pp. 57–60.
[41] F. Sadat, N. Habash, Combination of Arabic preprocessing schemes for statistical machine translation, in: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2006, pp. 1–8.
[42] S. Hasan, S. Mansour, H. Ney, A comparison of segmentation methods and extended lexicon models for Arabic statistical machine translation, Machine translation 26 (2012) 47–65.
[43] H. Al-Haj, A. Lavie, The impact of Arabic morphological segmentation on broad-coverage English-to-Arabic statistical machine translation, Machine translation 26 (2012) 3–24.
[44] I. Badr, R. Zbib, J. Glass, Segmentation for english-to-arabic statistical machine translation, in: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers, HLT-Short '08, Association for Computational Linguistics, Stroudsburg, PA, USA, 2008, pp. 153–156. URL: http://dl.acm.org/citation.cfm?id=1557690.1557732.
[45] A. El Kholy, N. Habash, Orthographic and morphological processing for English-Arabic statistical machine translation, Machine Translation 26 (2012) 25–45.
[46] B. Chen, M. Cettolo, M. Federico, Reordering rules for phrase-based statistical machine translation, in: International Workshop on Spoken Language Translation (IWSLT) 2006, 2006.
[47] N. Habash, Syntactic preprocessing for statistical machine translation, Proceedings of the 11th MT Summit 10 (2007).
[48] M. Carpuat, Y. Marton, N. Habash, Improving Arabic-to-English statistical machine translation by reordering post-verbal subjects for alignment, in: Proceedings of the ACL 2010 Conference Short Papers, Association for Computational Linguistics, 2010, pp. 178–183.
[49] M. Carpuat, Y. Marton, N. Habash, Improved Arabic-to-English statistical machine translation by reordering post-verbal subjects for word alignment, Machine Translation 26 (2012) 105–120.
[50] A. Bisazza, D. Pighin, M. Federico, Chunk-lattices for verb reordering in Arabic-English statistical machine translation, Machine translation 26 (2012) 85–103.
[51] J. Elming, N. Habash, Syntactic reordering for English-Arabic phrase-based machine translation, in: Proceedings of the EACL 2009 Workshop on Computational Approaches to Semitic Languages, Association for Computational Linguistics, 2009, pp. 69–77.
[52] J. Elming, Syntactic reordering integrated with phrase-based smt, in: Proceedings of the Second Workshop on Syntax and Structure in Statistical Translation, Association for Computational Linguistics, 2008, pp. 46–54.
[53] I. Badr, R. Zbib, J. Glass, Syntactic phrase reordering for english-to-arabic statistical machine translation, in: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, EACL '09, Association for Computational Linguistics, Stroudsburg, PA, USA, 2009, pp. 86–93. URL: http://dl.acm.org/citation.cfm?id=1609067.1609076.
[54] M. S. Hadj Ameur, A. Guessoum, F. Meziane, A pos-based preordering approach for english-to-arabic statistical machine translation, in: A. Lachkar, K. Bouzoubaa, A. Mazroui, A. Hamdani, A. Lekhouaja (Eds.), Arabic Language Processing: From Theory to Practice, Springer International Publishing, Cham, 2018, pp. 34–49.
[55] F. Sadat, E. Mohamed, Improved Arabic-French machine translation through preprocessing schemes and language analysis, in: Canadian Conference on Artificial Intelligence, Springer, 2013, pp. 308–314.
[56] E. Mohamed, F. Sadat, Hybrid Arabic-French machine translation using syntactic re-ordering and morphological preprocessing, Computer Speech & Language 32 (2015) 135–144.
[57] A. Alqudsi, N. Omar, K. Shaker, A hybrid rules and statistical method for arabic to english machine translation, in: 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), IEEE, 2019, pp. 1–7.
[58] A. Ittycheriah, S. Roukos, A maximum entropy word aligner for Arabic-English machine translation, in: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2005, pp. 89–96.
[59] V. Fossum, K. Knight, S. Abney, Using syntax to improve word alignment precision for syntax-based machine translation, in: Proceedings of the Third Workshop on Statistical Machine Translation, Association for Computational Linguistics, 2008, pp. 44–52.
[60] U. Hermjakob, Improved word alignment with statistics and linguistic heuristics, in: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1, Association for Computational Linguistics, 2009, pp. 229–237.
[61] Q. Gao, F. Guzman, S. Vogel, Emdc: a semi-supervised approach for word alignment, in: Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics, 2010, pp. 349–357.
[62] J. Riesa, D. Marcu, Hierarchical search for word alignment, in: Proceedings of the 48th annual meeting of the association for computational linguistics, Association for Computational Linguistics, 2010, pp. 157–166.
[63] I. T. Khemakhem, S. Jamoussi, A. B. Hamadou, Arabic-English semantic word class alignment to improve statistical machine translation, in: Proceedings of the International Conference Recent Advances in Natural Language Processing, 2015, pp. 663–671.
[64] M. Ellouze, W. Neifar, L. H. Belguith, Word alignment applied on english-arabic parallel corpus, in: LPKM, 2018.
[65] S. Berrichi, A. Mazroui, Benefits of morphosyntactic features on english-arabic statistical machine translation, in: 2018 IEEE 5th International Congress on Information Science and Technology (CiSt), IEEE, 2018, pp. 244–248.
[66] T. Brants, A. C. Popat, P. Xu, F. J. Och, J. Dean, Large language models in machine translation, in: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), 2007.
[67] S. Carter, C. Monz, Discriminative syntactic reranking for statistical machine translation, in: Ninth Conference of the Association for Machine Translation in the Americas (AMTA 2010), Denver, CO, USA, 2010.
[68] J. Niehues, T. Herrmann, S. Vogel, A. Waibel, Wider context by using bilingual language models in machine translation, in: Proceedings of the Sixth Workshop on Statistical Machine Translation, Association for Computational Linguistics, 2011, pp. 198–206.

[69] I. T. Khemakhem, S. Jamoussi, A. B. Hamadou, Integrating morpho-syntactic features in English-Arabic statistical machine translation, in: Proceedings of the Second Workshop on Hybrid Approaches to Translation, 2013, pp. 74–81.
[70] N. Habash, Four techniques for online handling of out-of-vocabulary words in Arabic-English statistical machine translation, in: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers, Association for Computational Linguistics, 2008, pp. 57–60.
[71] Y. Marton, D. Chiang, P. Resnik, Soft syntactic constraints for Arabic-English hierarchical phrase-based translation, Machine Translation 26 (2012) 137–157.
[72] K. Toutanova, H. Suzuki, A. Ruopp, Applying morphology generation models to machine translation, Proceedings of ACL-08: HLT (2008) 514–522.
[73] H. Sajjad, F. Dalvi, N. Durrani, A. Abdelali, Y. Belinkov, S. Vogel, Challenging language-dependent segmentation for arabic: An application to machine translation and part-of-speech tagging, arXiv preprint arXiv:1709.00616 (2017).
[74] M. Oudah, A. Almahairi, N. Habash, The impact of preprocessing on arabic-english statistical and neural machine translation, in: M. L. Forcada, A. Way, B. Haddow, R. Sennrich (Eds.), Proceedings of Machine Translation Summit XVII Volume 1: Research Track, MTSummit 2019, Dublin, Ireland, August 19-23, 2019, European Association for Machine Translation, 2019, pp. 214–221. URL: https://www.aclweb.org/anthology/W19-6621/.
[75] M. S. Hadj Ameur, A. Guessoum, F. Meziane, Improving arabic neural machine translation via n-best list reranking, Machine Translation 33 (2019) 279–314. URL: https://doi.org/10.1007/s10590-019-09237-6. doi:doi:10.1007/s10590-019-09237-6.
[76] F. Aqlan, X. Fan, A. Alqwbani, A. Al-Mansoub, Arabic–chinese neural machine translation: Romanized arabic as subword unit for arabic-sourced translation, IEEE Access 7 (2019) 133122–133135. doi:doi:10.1109/ACCESS.2019.2941161.
[77] S. Ding, A. Renduchintala, K. Duh, A call for prudent choice of subword merge operations in neural machine translation, in: Proceedings of Machine Translation Summit XVII Volume 1: Research Track, European Association for Machine Translation, Dublin, Ireland, 2019, pp. 204–213. URL: https://www.aclweb.org/anthology/W19-6620.
[78] D. Ataman, W. Aziz, A. Birch, A latent morphology model for open-vocabulary neural machine translation, in: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020, OpenReview.net, 2020. URL: https://openreview.net/forum?id=BJxSI1SKDH.
[79] D. Ataman, O. Firat, M. A. Di Gangi, M. Federico, A. Birch, On the importance of word boundaries in character-level neural machine translation, in: Proceedings of the 3rd Workshop on Neural Generation and Translation, 2019, pp. 187–193.
[80] X. Liu, D. F. Wong, Y. Liu, L. S. Chao, T. Xiao, J. Zhu, Shared-private bilingual word embeddings for neural machine translation, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 3613–3622.
[81] P. Shapiro, K. Duh, Morphological word embeddings for Arabic neural machine translation in low-resource settings, in: Proceedings of the Second Workshop on Subword/Character LEvel Models, 2018, pp. 1–11.
[82] Y. Belinkov, N. Durrani, F. Dalvi, H. Sajjad, J. Glass, What do neural machine translation models learn about morphology?, arXiv preprint arXiv:1704.03471 (2017).
[83] D. Ataman, M. Federico, An evaluation of two vocabulary reduction methods for neural machine translation, in: Proceedings of the The 13th Conference of The Association for Machine Translation in the Americas, Boston, USA, 2018, pp. 97–110.
[84] M. García-Martínez, W. Aransa, F. Bougares, L. Barrault, Addressing data sparsity for neural machine translation between morphologically rich languages, Machine Translation 34 (2020) 1–20. URL: https://doi.org/10.1007/s10590-019-09242-9. doi:doi:10.1007/s10590-019-09242-9.
[85] Y. Nishimura, K. Sudoh, G. Neubig, S. Nakamura, Multi-source neural machine translation with missing data, IEEE/ACM Transactions on Audio, Speech, and Language Processing 28 (2019) 569–580.
[86] X. Tan, J. Chen, D. He, Y. Xia, Q. Tao, T.-Y. Liu, Multilingual neural machine translation with language clustering, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 962–972.
[87] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, L. Zettlemoyer, Multilingual denoising pre-training for neural machine translation, arXiv (2020) arXiv–2001.
[88] R. Aharoni, M. Johnson, O. Firat, Massively multilingual neural machine translation, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 3874–3884.
[89] E. H. Almansor, A. Al-Ani, A hybrid neural machine translation technique for translating low resource languages, in: International Conference on Machine Learning and Data Mining in Pattern Recognition, Springer, 2018, pp. 347–356.
[90] B. Ji, Z. Zhang, X. Duan, M. Zhang, B. Chen, W. Luo, Cross-lingual pre-training based transfer for zero-shot neural machine translation, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, 2020, pp. 115–122.
[91] A. Almahairi, K. Cho, N. Habash, A. Courville, First result on Arabic neural machine translation, arXiv preprint arXiv:1606.02680 (2016).
[92] M. Junczys-Dowmunt, T. Dwojak, H. Hoang, Is neural machine translation ready for deployment, A case study on 30 (2016).
[93] Y. Belinkov, J. Glass, Large-scale machine translation between Arabic and Hebrew: Available corpora and initial results, arXiv preprint arXiv:1609.07701 (2016).
[94] Y. Salem, A. Hensman, B. Nolan, Implementing Arabic-to-English machine translation using the role and reference grammar linguistic model, in: In Proceedings of the Eighth Annual International Conference on Information Technology and Telecommunication (ITT 2008), At Ireland, Dublin Institute of Technology, 2008.
[95] O. Shirko, N. Omar, H. Arshad, M. Albared, Machine translation of noun phrases from Arabic to English using transfer-based approach, Journal of Computer Science 6 (2010) 350.
[96] A. Soudi, V. Cavalli-Sforza, A. Jamari, A prototype English-to-Arabic interlingua-based mt system, in: Proceedings of the third international conference on language resources and evaluation: workshop on Arabic language resources and evaluation: status and prospects. Las Palmas de Gran Canaria, Spain, 2002, pp. 18–25.
[97] K. Shaalan, A. Rafea, A. A. Moneim, H. Baraka, Machine translation of English noun phrases into Arabic, International Journal of Computer Processing of Oriental Languages 17 (2004) 121–134.

[98] R. Al Dam, A. Guessoum, Building a neural network-based english-to-arabic transfer module from an unrestricted domain, in: Machine and Web Intelligence (ICMWI), 2010 International Conference on, IEEE, 2010, pp. 94–101.

[99] L. Hadla, T. Hailat, M. Al-Kabi, Evaluating arabic to english machine translation, International Journal of Advanced Computer Science and Applications 5 (2014) 68–73.

[100] A. Guessoum, R. Zantout, A methodology for a semi-automatic evaluation of the lexicons of machine translation systems, Machine translation 16 (2001) 127–149.

[101] A. Guessoum, R. Zantout, A methodology for evaluating arabic machine translation systems, Machine Translation 18 (2004) 299–335.

[102] N. Adly, S. Al Ansary, Evaluation of Arabic machine translation system based on the universal networking language, in: International Conference on Application of Natural Language to Information Systems, Springer, 2009, pp. 243–257.

[103] A. Al-Rukban, A. K. J. Saudagar, Evaluation of English to Arabic machine translation systems using bleu and gtm, in: Proceedings of the 2017 9th International Conference on Education Technology and Computers, ACM, 2017, pp. 228–232.

[104] F. Guzmán, H. Bouamor, R. Baly, N. Habash, Machine translation evaluation for Arabic using morphologically-enriched embeddings, in: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016, pp. 1398–1408.

[105] M. El Marouani, T. Boudaa, N. Enneya, Incorporation of linguistic features in machine translation evaluation of Arabic, in: International Conference on Big Data, Cloud and Applications, Springer, 2018, pp. 500–511.

[106] N. Habash, O. Rambow, R. Roth, Mada+ tokan: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, pos tagging, stemming and lemmatization, in: Proceedings of the 2nd international conference on Arabic language resources and tools (MEDAR), Cairo, Egypt, volume 41, 2009, p. 62.

[107] R. C. Moore, Improving ibm word-alignment model 1, in: Proceedings of the 42nd annual meeting on association for computational linguistics, Association for Computational Linguistics, 2004, p. 518.

[108] S. Vogel, H. Ney, C. Tillmann, Hmm-based word alignment in statistical translation, in: Proceedings of the 16th conference on Computational linguistics-Volume 2, Association for Computational Linguistics, 1996, pp. 836–841.

[109] R. Sennrich, B. Haddow, A. Birch, Neural machine translation of rare words with subword units, arXiv preprint arXiv:1508.07909 (2015).

[110] W. Ling, I. Trancoso, C. Dyer, A. W. Black, Character-based neural machine translation, arXiv preprint arXiv:1511.04586 (2015).

[111] Y. Kim, Y. Jernite, D. Sontag, A. M. Rush, Character-aware neural language models., in: AAAI, 2016, pp. 2741–2749.

[112] J. Sun, B. Feng, W. Xu, Particle swarm optimization with particles having quantum behavior, in: Proceedings of the 2004 congress on evolutionary computation (IEEE Cat. No. 04TH8753), volume 1, IEEE, 2004, pp. 325–331.

[113] D. Ataman, M. Negri, M. Turchi, M. Federico, Linguistically motivated vocabulary reduction for neural machine translation from turkish to english, The Prague Bulletin of Mathematical Linguistics 108 (2017) 331–342.

[114] R. Sennrich, B. Haddow, Linguistic input features improve neural machine translation, in: Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers, Association for Computational Linguistics, Berlin, Germany, 2016, pp. 83–91. URL: `https://www.aclweb.org/anthology/W16-2209`. doi:doi:10.18653/v1/W16-2209.

[115] F. Burlot, M. García-Martínez, L. Barrault, F. Bougares, F. Yvon, Word representations in factored neural machine translation, in: Proceedings of the Second Conference on Machine Translation, Association for Computational Linguistics, Copenhagen, Denmark, 2017, pp. 20–31. URL: `https://www.aclweb.org/anthology/W17-4703`. doi:doi:10.18653/v1/W17-4703.

[116] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 7871–7880. URL: `https://www.aclweb.org/anthology/2020.acl-main.703`.

[117] M. Ziemski, M. Junczys-Dowmunt, B. Pouliquen, The united nations parallel corpus v1.0, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), European Language Resources Association (ELRA), Portorož, Slovenia, 2016, pp. 3530–3534. URL: `https://www.aclweb.org/anthology/L16-1561`.

[118] P. Lison, J. Tiedemann, Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles (2016).

[119] N. Habash, N. Zalmout, D. Taji, H. Hoang, M. Alzate, A parallel corpus for evaluating machine translation between Arabic and European languages, in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, Association for Computational Linguistics, Valencia, Spain, 2017, pp. 235–241. URL: `https://www.aclweb.org/anthology/E17-2038`.

[120] M. Cettolo, C. Girardi, M. Federico, Wit^3: Web inventory of transcribed and translated talks, in: Proceedings of the 16^{th} Conference of the European Association for Machine Translation (EAMT), Trento, Italy, 2012, pp. 261–268.

[121] T. Zerrouki, A. Balla, Tashkeela: Novel corpus of arabic vocalized texts, data for auto-diacritization systems, Data in brief 11 (2017) 147.

[122] M. Alrabiah, A. Alsalman, E. Atwell, The design and construction of the 50 million words ksucca king saud university corpus of classical arabic., 2013.

[123] S. Alansary, M. Nagi, The international corpus of arabic: Compilation, analysis and evaluation, in: Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), 2014, pp. 8–17.

[124] M. Abbas, K. Smaili, Comparison of topic identification methods for arabic language, in: Proceedings of International Conference on Recent Advances in Natural Language Processing, RANLP, 2005, pp. 14–17.

[125] M. Abbas, K. Smaïli, D. Berkani, Evaluation of topic identification methods on arabic corpora., J. Digit. Inf. Manag. 9 (2011) 185–192.

[126] M. Attia, M. Yaseen, K. Choukri, Specifications of the arabic written corpus produced within the nemlar project, 2005.

[127] M. Boudchiche, A. Mazroui, Enrichment of the nemlar corpus by the lemma tag, in: Workshop Language Resources of Arabic NLP: Construction, Standardization, Management and Exploitation. Rabat, Morocco, 2015.

[128] G. Klein, Y. Kim, Y. Deng, V. Nguyen, J. Senellart, A. Rush, OpenNMT: Neural machine translation toolkit, in: Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers), Association for Machine Translation in the Americas, Boston, MA, 2018, pp. 177–184. URL: `https://www.aclweb.org/`

anthology/W18-1817.

[129] M. Ott, S. Edunov, A. Baevski, A. Fan, S. Gross, N. Ng, D. Grangier, M. Auli, fairseq: A fast, extensible toolkit for sequence modeling, in: Proceedings of NAACL-HLT 2019: Demonstrations, 2019.

[130] A. Vaswani, S. Bengio, E. Brevdo, F. Chollet, A. N. Gomez, S. Gouws, L. Jones, L. Kaiser, N. Kalchbrenner, N. Parmar, R. Sepassi, N. Shazeer, J. Uszkoreit, Tensor2tensor for neural machine translation, CoRR abs/1803.07416 (2018). URL: http://arxiv.org/abs/1803.07416.

[131] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, et al., Moses: Open source toolkit for statistical machine translation, in: Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions, Association for Computational Linguistics, 2007, pp. 177–180.

[132] S. Green, D. Cer, C. D. Manning, Phrasal: A toolkit for new directions in statistical machine translation, in: In Proceddings of the Ninth Workshop on Statistical Machine Translation, 2014.

[133] R. Sennrich, B. Haddow, A. Birch, Neural machine translation of rare words with subword units, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Berlin, Germany, 2016, pp. 1715–1725. URL: https://www.aclweb.org/anthology/P16-1162. doi:doi:10.18653/v1/P16-1162.

[134] F. J. Och, H. Ney, A systematic comparison of various statistical alignment models, Computational Linguistics 29 (2003) 19–51.

[135] S. L. Smith, D. H. Turban, S. Hamblin, N. Y. Hammerla, Offline bilingual word vectors, orthogonal transformations and the inverted softmax, arXiv preprint arXiv:1702.03859 (2017).

[136] R. Řehůřek, P. Sojka, Software Framework for Topic Modelling with Large Corpora, in: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, ELRA, Valletta, Malta, 2010, pp. 45–50. http://is.muni.cz/publication/884893/en.

[137] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information, Transactions of the Association for Computational Linguistics 5 (2017) 135–146.

[138] S. Sharma, L. El Asri, H. Schulz, J. Zumer, Relevance of unsupervised metrics in task-oriented dialogue for evaluating natural language generation, CoRR abs/1706.09799 (2017). URL: http://arxiv.org/abs/1706.09799.