RESEARCH ARTICLE OPEN ACCESS

Divergence and Ambiguity Control in an English to Arabic Machine Translation

Marwan Akeel*, R. B. Mishra**
*(Department of Computer Engineering, Indian Institute of Technology-(BHU), Varanasi)
** (Department of Computer Engineering, Indian Institute of Technology (BHU), Varanasi)

ABSTRACT
The aim of the translation is to transfer a meaning from one language to another. This process requires the knowledge of both source and target languages linguistic features. Human languages are highly ambiguous, and differently in different languages. In this paper, we discuss the language divergences and ambiguities exist in English into Arabic machine translation and the control methods we have used in our developed ANN and rule based machine translation system to tackle these problems in order to minimize the error rate and have better translation.
Keywords – ANN, English Arabic translation, machine translation, neural network, NLP, rule-based

I. INTRODUCTION

Machine Translation is the process by which computer software is used to translate a text from one natural language into another language with or without minimal human intervention. This definition involves accounting for the grammatical structure of each language and using rules and grammars to transfer the grammatical structure of the source language into the target language. There are three type of machine translation. Machine aided human translation, human aided machine translation and fully automated machine translation. Here, we are focusing on the fully automated machine translation. In general, there are two steps that the input of a source language, i.e. the text to be translated, to the machine translation system is gone through to produce the output of the target language. First, morphologically and syntactically analysing of the source text. Second, replacing the source text with the appropriate target language text or meaning.

The transferred meaning and semantics is the most significant point of focus. It requires our utmost attention of the problems that may affect the meaning. The problems are mainly due to the differences in linguistic systems and languages. As put by [1] "Errors and problems in translation mostly result from the non-equivalence between the source and target languages". [2] defines translation problems as "a linguistic element that becomes a translation problem when the translator has to decide between more than one way of rendering it". Ambiguity in main aspect is the property of words, terms, notations, signs, symbols, and concepts (within a particular context) as being undefined, indefinable, multi-defined, or without an obvious definition, and thus having a misleading, or unclear, meaning. Ambiguity could be a result of a language divergence. Reducing the errors during analysing and selecting of meanings will improve the output of the translation.

Many works have been done to reduce the differences gap existed in English to Arabic machine translation. [3] focuses on the specific challenges of machine translation with Arabic either as source or target language. [4] discussed modern standard Arabic and machine translation issues about Arabic, such as morphology and Arabic script. [5] focuses on word agreement and ordering in a ruled based English-Arabic machine translation. [6] investigates different methodologies to manage the problem of morphological and syntactic ambiguities in Arabic with a view to machine translation. [7] achieved significant improvements in translation quality after extending a pre-translation syntactic reordering approach developed on a close language pair (English-Danish) to the distant language pair, English-Arabic in a statistical machine translation.

In this paper we are going to discuss the most factors impacting the transfer of the source text meaning correctly. These factors are mainly language divergences and ambiguities. In next section, we are given a preview of our developed ANN and Rule based English into Arabic machine translation system [8].

In section three, we discuss some of the ambiguity and divergences problems occurring in English into Arabic machine translation and mentioned the control methods used to resolve it. Finally, the implementation of the methods used to control the ambiguity and divergence problems in our system.

Arabic transliteration are provided in Buckwalter transliteration scheme [9].

II. SYSTEM ARCHITECTURE

We have implemented English into Arabic Machine Translation system for translating simple well-structured English sentences into well-structured Arabic sentences using a rule-based approach, feed-forward back-propagation artificial neural network (ANN). Our system is able to translate sentences having gerunds, infinitives, prepositions and prepositional objects, direct objects, indirect objects, etc. The neural network works as a bilingual dictionary. The bilingual dictionary matches an English word to one or sometimes two meanings in Arabic. The bilingual dictionary does not only store the meaning of English word in Arabic but it also stores linguistic features attached to the word. The block diagram of our English to Arabic Machine Translation System is shown in Fig 1.

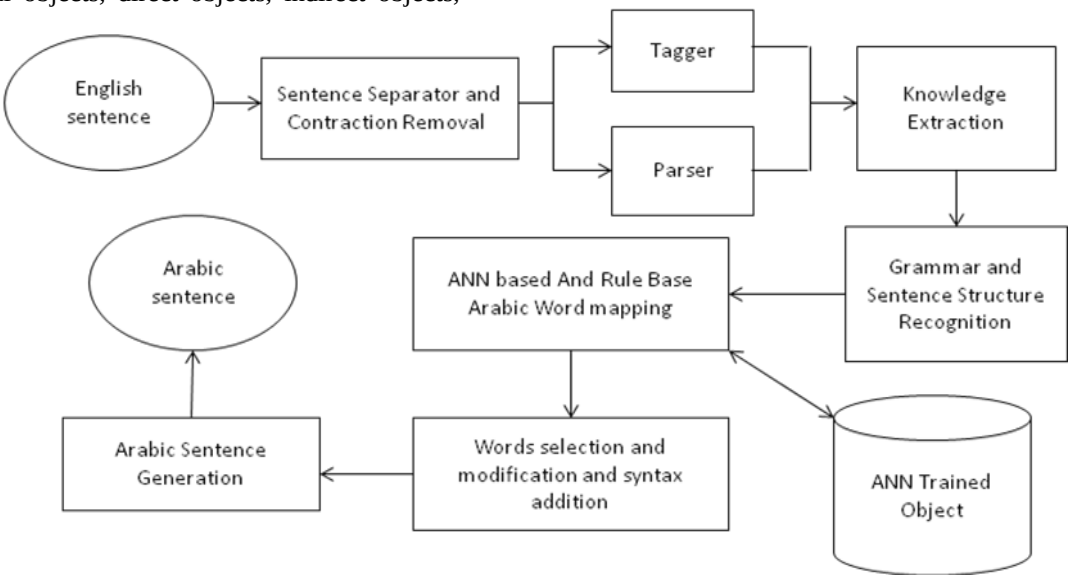

Fig 1: System architecture

There are eight main modules on our system. The modules and the function of each module are explained below:

Sentence Separator and Contractions Removal: This module first separates the paragraph into sentences. Then each sentence is processed. If any contraction is present in the sentence, it is removed.

Parser and Tagger**:** Stanford typed dependency parser is used for parsing the English Text [10]. The statistical parser derives the most probable parse of a sentence and the grammatical relations among the sentence words. The dependency is written as abbreviated_relation_name(governor, dependent). We also used Stanford POS tagger for tagging the English text[11]. A Part-Of-Speech Tagger (POS Tagger) assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc.

Knowledge Extraction: This module extracts information from parser and tagger for each part of the sentence. Each part of the sentence is converted to knowledgeable object by adding all the information associated with it and sentence is represented as a collection of knowledgeable objects.

Grammar Analysis and Sentence Structure Recognition: This module processes the collection of knowledgeable objects and recognizes parts of the sentence e.g. subject, main verb, auxiliary verb, object, indirect object etc. Each recognized part of the sentence contains one or more word called chunk. So, each sentence is divided into chunks. The grammatical attributes of the sentence such as tense, voice, type, and form are also identified. This module generates the grammatical structure of the English sentence by analysing the chunks and attributes of the sentence. It also decides the Arabic sentence structure and word order according to the English recognized sentence structure.

ANN and Rule based mapping: The English POS or chunks recognized at previous module are sent to the ANN module. English sentence chunks (or phrases) are processed according to the Arabic translation structure for translation. Each English word of a chunk is encoded into numeric form by the Encoder-Decoder. The ANN module looks for each word in the bilingual dictionary object which is trained for word mapping, and gets the corresponding Arabic word and associated information in numeric form. This result is decoded to textual form by the Encoder-Decoder and sent back to the main module.

Words Selection and Modification and Syntax Addition: In this module, we select the correct form out of the received Arabic translation of each chunk word. The selection is done depending on the sentence tense or the features of the chunk main word or the subject main word. Then the necessary adaptation is done by adding/deleting affixes to/from the selected form or word.

Arabic Sentence Generation: In this Module, the chunks are arranged in accordance with the Arabic sentence structure obtained previously and the necessary additional words are added.

ANN Module: We used feed-forward back-propagation artificial neural network for the selection of Arabic words/tokens (such as verb, noun/pronoun etc) equivalent to English words/tokens. Each English word is matched to at least two meanings in Arabic.

III. ARABIC ENGLISH LANGUAGE AMBIGUITY AND DIVERGENCE

Divergence is a language dependent phenomenon. It is not necessary that same set of divergences will occur across all the languages. Translation ambiguity occurs when a word in one language can be translated in more than one way into another language. Language divergence in many cases causes translation ambiguity. We are going to discuss most of the commonly found divergences and ambiguities in English to Arabic MT, the lexical and syntactic ambiguity, and the structural divergences.

3.1 Lexical Ambiguity

Lexical ambiguity arises when there are potentially two or more ways in which a word can be analysed. In other words, lexical ambiguities are where one word can be interpreted in more than one way. [12] establish three basic types of lexical ambiguity: category ambiguity, homographs ambiguity and transfer or translational ambiguity. In addition to the three basic types, we also add the pronoun reference ambiguity, and number and gender ambiguity as separate special types. At the following subsections, we discuss the three main types of the lexical ambiguity and the control methods applied to control it. Then separately discuss the two special types and the control methods designed to encounter them.

3.1.1 Category ambiguity

Category ambiguity is the simplest type of lexical ambiguity where a given word could be assigned to more than one grammatical or syntactic category according to the context. Examples of this type in English: words like clean, free and plain can be a noun, verb or adjective. Other words like watch, book and control can be a noun or verb. In Arabic there are some words that can be in more than one category, for example يزيد (yzyd) could be a verb with meaning of "increase", or a person name. the word على (ElY) can come as a preposition with meaning of "on" or a verb with meaning of "raise", consider the following example:

1) He is calculating.
   إنه يحسب (<nh yHsb)
   إنه أناني (<nh >nAny)

In the example, the word "calculation" falls in two category. It can be a verb with meaning of "working out by using numbers", or it can also function as an adjective depicting a characteristic of a person with meaning of "selfish, scheming, or shrewd".

This type of ambiguity is mostly solved by morphological inflection. Syntactic parsing solutions are also possible.

3.1.2 Homograph ambiguity

This type of ambiguity occurs when a word can have two or more different meanings. Homographs are a pair (or a group) of words that are spelled the same way, but have different meanings. They may or may not be pronounced the same way. An English example is the word "row" which could be line, argument, or propel a boat. Also, the word "bat" which means piece of sports equipment or an animal. Many Arabic words have more than one meaning like مركز (mrkz) which can means center, position, rank or status. Also, موقع (mwqE) could be position, rank, site or status. Consider the following examples:

2) I went to the bank.
   ذهبت إلى المصرف. (*hbt <lY AlmSrf)
   ذهبت إلى ضفة النهر. (*hbt <lY Dfp Alnhr)
3) She can not bear children.
   لاتستطيع تحمل الأطفال. (lA tstTyE jHml Al>TfAl)
   لا تستطيع إنجاب الأطفال. (lA tstTyE <njAb Al>TfAl)

The meaning of the word "bank" in example 2 is ambiguous where it can mean "establishment for keeping money, valuables", or "a land sloping up along side of a river or canal". The word "bear" in example 3 also has two different meaning, first meaning is "afford or endure something" and the second meaning is "give birth to".

To figure out the exact meanings of such words, it demands a careful consideration of the specific context. In our system, the ANN bilingual data base contains only one Arabic meaning of an English word and sometimes two meanings if the English word can come as a noun or verb.

3.1.3 Transfer or translational ambiguity

This type arises when a single source language word can potentially be translated into a number of different target language words or expressions. The source language by itself is not ambiguous, it is just ambiguous perspective of the target language. Most of the English words have more than one translation in Arabic. For example:

4) My friend helps the old man.
   يساعد صديقي الرجل العجوز (ysAEd Sdyqy Alrjl AlEjwz)
   يساعد صديقي الرجل القديم (ysAEd Sdyqy Alrjl Alqdym)

When a machine translation translates the previous example, the possibility of not selecting the appropriate Arabic translation of the word "old", which is Compatible with the word "man", is present since "old" has different forms of the similar meaning in Arabic depending on the context. The machine translation maybe translate the word "old"

into (someone old or elderly) which is the correct one, or into (something old or outdated) which is wrong.

Control method of the basic types of lexical ambiguity: The problem of bilingual lexical ambiguity is solved by contextual rules (Rule Based Machine Translation), or examples (Example Based Machine Translation), or frequencies and 'language models' (Statistical Machine Translation) [13]. In our System, we used a statistical parser to derive the most probable parse of a sentence and the grammatical relations among the sentence words. In addition to the parser we support the system with a POS tagger for word-category disambiguation where the POS tagger assigns parts of speech tag to each word and tokens such as nouns, verbs, adjectives etc. We also designed an ANN one to one dictionary which maps each English word to at most two meanings in Arabic. English words which have two different meanings from different category are mapped into two Arabic meanings. We have to mention here that an Arabic word is a noun, a verb or an article. A noun in Arabic is equivalent to what we know in English as nouns, pronouns, adjectives, and adverbs. For example, the English word "book" is mapped in the bilingual dictionary into:

"كتاب : ن8 حجز7 محجوز6 يحجز5 حجز4 أحجز3 يحجز2 حجز1 |tp #sing *masc"

The first part of the translation contains the seven forms of the Arabic verb plus the flag value while the second part contains the Arabic noun meaning and the noun features. The two meanings are separated by a colon (:).

The three techniques (parser, tagger, two different categorical meanings for each word) were applied to tackle the three types of the lexical Ambiguity. The pseudo-code for selecting the right translation of a word which has verb and noun translations:

Analyse the sentence using the Parser and Tagger
Attach to each word its POS tag and the grammatical relation it participate in
Divide the sentence into chunks
Rearrange the order of the sentence chunks according to the Arabic generated word order
For all the chunks
  Translate the chunk word by word starting from the right most word
  Translate the word
  If the retrieved translation contains (:) Then
    If the word POS tag is noun then Select the second half starting from the (:)
    Else Select the first half up to the (:)

3.1.4 Pronoun reference ambiguity

A pronoun is a word that takes the place of a noun. An ambiguous pronoun reference occurs when it is not clear what noun a pronoun refers to. Arabic has same pronouns like English but with more specification for the pronouns "you", "they" and "it". Second person pronouns vary in number and gender in Arabic while in English has only one form. The same problem also with the third person plural pronouns "they". Example, matching the pronoun "you" to an Arabic pronoun is not an easy task since the gender of the intended you and how many person related to the pronoun should be recognized. Similar problems occur also with the possession pronouns and object pronouns. Table 1 shows the English pronouns "you" and "they" and their corresponding Arabic pronouns.

Table 1: The English pronouns you and they and the corresponding Arabic pronouns.

English	Number & gender	Arabic
You	Singular masc.	أنت (>nt)
	Singular fem.	أنتِ (>nti)
	Dual masc.	أنتما (>ntmA)
	Dual fem.	أنتما (>ntmA)
	Plural masc.	أنتم (>ntm)
	Plural fem.	أنتن (>ntn)
They	Dual masc.	هما (hmA)
	Dual fem.	هما (hmA)
	Plural masc.	هم (hm)
	Plural fem.	هن (hn)

Control method: to tackle the ambiguous pronoun references problem, we created two global variables to save the person, number and gender word features of the last occurring subject and object. These global variables are set at a sentence and used when analysing the second sentence subject. The global variable value should match the ambiguous pronoun in the person feature and if necessary the number and gender features too in order to be used. Having two global variables increases the chance of clarifying any pronoun ambiguity. Following example illustrate the benefit of global variable:

5) Ahmed met his parents. They were very tired.

قابل أحمد والديه. كانا متعبين جدا. (qAbl >Hmd wAldyh. kAnA mtEbyn jdA)

In example 5, we have two global variables carrying the features of the words "Ahmed" and "parents". First global value carries the word features of the subject and the second global variable carries the word features of the object. When analysing the subject pronoun "they" at the second sentence we find that it is one of the ambiguous pronouns. To clarify the subject pronoun features we compare its default word features (plural, 3rd person, masculine) with the first global variable value (singular, 3rd person, masculine) and the second global variable value (dual, 3rd person, masculine). The first global value has a conflict value with the pronoun subject since they differ in the number feature (singular and dual). It is found that the pronoun subject features are closer to the word features of the second global variable (where dual is considered part of the plural in English). So the subject pronoun "they" is given the features (dual, 3rd person, masculine) which are the features of the word "parents". Following is the

pseudo-code to detect the pronoun subject word features for the pronoun "they" and "you". Other pronoun ambiguous cases are solved following similar steps.

If the subject is pronoun and equal to "you" or "they" or "it"
  Compare the pronoun subject default word features with the first global variable and second global variable
  Select the appropriate match.
  If both global variables have conflict values with subject pronoun Then
    Use the pronoun default features and reset the related global variable to the new values.

3.1.5 Gender and number ambiguity

A system of grammatical gender, whereby every noun was treated as either masculine, feminine or neuter, existed in old English. Gender is no longer an inflectional category in modern English. The only traces of the old English gender system are found in the system of pronoun. Pronouns such as he and she are used to refer specifically to persons or animals of one or other gender, and certain others (such as it) for sexless objects. Another manifestation of natural gender that continues to function in English is the use of certain nouns to refer specifically to persons or animals of a particular gender. In English, The verb and adjective form do not differentiate between male and female. The Arabic language on the other hand pays more attention to gender since it determines the selection of grammatical forms. Arabic nouns are either marked for masculine or feminine which is also relevant to pronouns, verbs and adjectives. Usually when referring to a male, a masculine noun is used and when referring to a female, a feminine noun is used. It's not just nouns referring to people that have gender. Inanimate objects (chairs, keys, bikes, etc.) is either masculine or feminine.

English like most world languages is distinguishing only between singular and plural number. Only the nouns and pronouns are marked for number. When speaking about one person or one thing, a singular form is used while the plural form is used when we speak about more than one. Arabic has singular, dual and plural forms of pronouns, nouns, verbs, adjectives, etc. The singular form is used when referring to one person or thing, the dual subject pronoun to refer to two people or two things, and the plural form when referring to more than two people or things.

In Arabic, the gender and number ambiguity affect not only the related word translation but also all other wards that have a formal agreement with it. For example:

6) That student is clever.
   Masculine: ذلك الطالب ذكي. (*lk AlTAlb *ky)
   That (singular-masculine) the student (singular-masculine) clever (singular-masculine)
   Feminine: تلك الطالبة ذكية. (tlk AlTAlbp *kyp)
   That (singular-feminine) the student (singular- feminine) clever (singular-feminine)

On the previous example, there are two possible Arabic translations demonstrated in two sentences. The two Arabic sentences differ from each other in the sentence subject where they are not having the same gender. Because of the agreement feature in Arabic the subject gender affects the subject complement and its adjective. On the other hand, the same English words are used for both genders.

7) They visited the two workers
   زاروا العاملين (zArwA AlEmlyn)
   Visited (plural-masculine) the workers (dual-masculine)

The above sentence is an example of the use of the grammatical number in English and Arabic. The subject pronoun "they" and the word "workers" are plural but the adjective "two" specifies the worker number to dual. The Arabic translation has the plural feature on زاروا (zArwA) which is a combination word containing the subject and the verb. The word العاملين (AlEmlyn) is the dual form of عامل (EAml) in addition to the definite article.

Control method: we used two methods to target the gender and number ambiguity. The first method is supporting the bilingual dictionary entries with the words features. Each noun retrieved is attached with its word features, for example: the English noun "table" is mapped to "طاولة |tp #sing *fem" where طاوله (TAwlp) is the Arabic translation and rest are the word features. The adjective "fast" is mapped to "سريع1 سريعة2 سريعان3 سريعتان4 سريعون5 سريعات6" where the six forms are two for the singular masculine and feminine, two for the dual masculine and feminine and remaining two forms are for the plural masculine and feminine.

The second method is the use of variables to keep any word features which may affect any coming word formation. From sentence to sentence we need global variables as we have shown on the previous subsection. The global variables are mainly to set the gender and number word features of any ambiguous pronoun. For a sentence, we use two local variables to save the subject and object main word features in case we need them in any verb subject relation or adjective noun relation. While translating a sentence chunk/phrase we also save the word features of the chunk main word/noun in a variable since it affects the formation of all other chunk words. Following is the pseudo-code to translate subject or object chunk which contains more than one word.

If the chunk contains "the" or possessive pronoun
Then Set definiteness = true
Translate the chunk word by word starting from the most right word
  Translate the word
  If it is the first word Then

Extract and Save the word features in the (temporary variable) and If it is a subject/object chunk Save the word features in the (subject/object local variable) and reset the (first/second global variable)
Do the necessary adjustment to the translation according to the word grammatical state
If the chunk contains possessive pronoun Then
If the possessive pronoun is ambiguous Then
Use the appropriate global variable to adjust the possessive pronoun word features
Add the appropriate possessive suffix to the Arabic translation word
If the definiteness = true Then Add ال to the beginning of the Arabic translation word
Add the word to the chunk translation variable
Else select the appropriate form of the Arabic translation forms according to the temporary variable value
Do the necessary adjustment to the Arabic translation word according to the word grammatical state
If definiteness = true Then Add ال to the beginning of the Arabic translation word
Add the word to the chunk translation variable

3.2 Syntactic ambiguity

Syntactically, sentences are made up of phrases; phrases are made up of words. Thus, there exists a kind of syntactic relation, i.e., a hierarchical relation. Such a relation shows the inner layers of sentences. The way the words are arranged hierarchically has an important role in determining the meaning. It is also because of this that the ambiguity is produced. Syntactic ambiguity arises when there is more than one way of analysing the underlying structure of a sentence according to the grammar used in the system. The following examples demonstrate this type of ambiguity:

8) Old men and women live in this complex
9) John mentioned the letter I sent to Sami.

In example 8, it is ambiguous between whether the men and the women are old or just the men. Also example 9 can be understood in two ways: The first one, John mentioned the letter which I sent to Sami. The second way, John mentioned to Sami the letter which I sent. Here, the syntax which is not clear not the meaning of the words. For the purposes of this discussion, we represent example 8 in the notation of a syntax-tree diagram.

The two trees in Fig 2 and Fig 3 represent under study parts of the two different analyses of sentence on the first example in the sense of recording two different 'parse histories'. In linguistic terms, they correspond to the two readings of the sentence: one in which the CC is at upper level of both nouns (men and women)and the ADJ, and the ADJ is in sibling location with the "men" word indicating that the Adjective "old" notate only the men. The other reading, the ADJ is one level upper than the CC and both nouns, and both nouns and the CC are siblings indicating that the adjective is related with the both men and women.

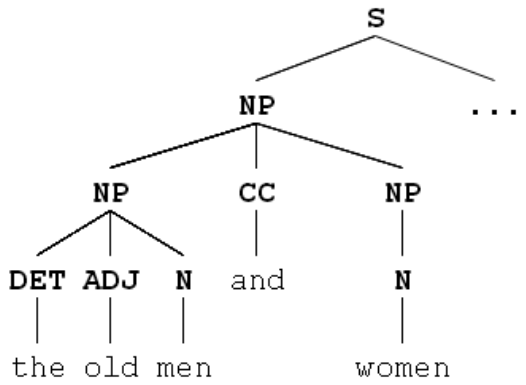

Fig 2: The syntax tree of the case when the adjective old only for the word men.

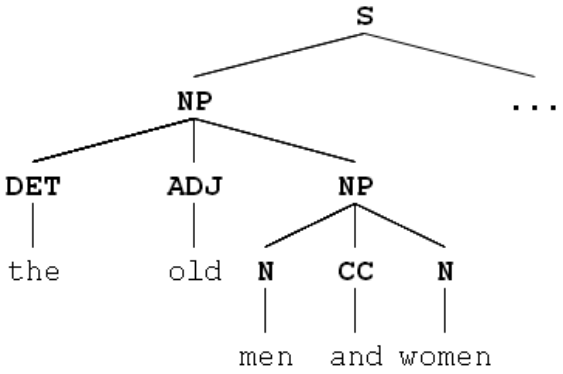

Fig 3: The syntax tree of the case when the the adjective old is for both men and women.

Control method: We used both the information of the grammatical relations among the sentence words generated by the statistical parser, and the assigned parts of speech to each word done by the POS tagger to tackle the syntactic ambiguity. Both tagger and parser still make some mistakes, but commonly work rather well.

3.3 Structural differences

In this subsection, we are going to discuss some English Arabic differences at a level upper than one word. These include global and local word order divergence, tense divergence and the word agreement divergence.

3.3.1 Word order divergence

Compared with many other languages, Subject (S) Verb (V) Object (O) word order in English is fairly rigid. Arabic is a relatively free word order language, structured under the combinations of SVO, VSO, VOS and OVS. The most common synchronize structures are SVO and VSO [14]. Following are an English sentence with one Arabic translation but with different word order.

10) Sami took the book.
أخذ سامي الكتاب (>x* sAmy AlktAb)
(VSO) Took Sami the book.
سامي أخذ الكتاب (sAmy >x* AlktAb)
(SVO) Sami took the book.

Another word order difference is present between the adjectives and the noun they modify. In English, the adjective precedes its noun while in Arabic adjectives always follow the noun they modify (see example 11).

11) A small room.
غرفة صغيرة (grfp Sgyrp)
room small.

The possessive pronoun/noun also precedes the possessor or possessors in English. In Arabic, possession is indicated by a noun which comes after what it owns, or by appending a suffix to a noun. This suffix reflects the gender and number features of the possessor or possessors. The possessive pronoun (its) does not exist in Arabic, it is referred to as (his) or (her) depending whether the noun is masculine or feminine. Examples 12 and 13 demonstrate word order divergence of both the possessive pronoun/noun and its possessor(s), and the adjective and its noun.

12) Their seat tickets.
تذاكر مقاعدهم (t*Akr mqAEdhm)
Tickets seat their.

13) Adam's car is new.
سيارة آدم جديدة (syArp Idm jdydp)
Car Adam new.

Control method: we selected two Arabic word order structures to be used on building the Arabic translation sentences. The two word order structures are SVO and VSO. The SVO structure is used when:

The sentence is nominal.
The sentence starts with the word إن (<n).
The sentence starts with the word كان (kAn).

All other sentences are in VSO word order. These word order structures are applied in all type of sentences.

For the reordering divergence, the sentence is divided into small parts called chunks. The translation starts first by reordering the chunks according to the Arabic final sentence structure. Then every chunk is translated separately one by one. Each chunk is translated starting from the right word in the chunk. In this way we reorder the chunk words and the output is the right Arabic word order. Following is the **pseudo-code** to translate a sentence.

Reorder the sentence chunks according to the Arabic sentence structure
For all the chunks Process one by one
  For every chunk
    Translate the words starting from the most right word
    Select the proper word translation
    Modify the selected translation word if needed
    Add the translation to the local chunk translation
    Add the chunk output translation to the sentence output translation
Add any necessary fix words
Output the translation

3.3.2 Tense and aspect divergence

There are 12 tenses in English which have resulted from the combination of both tense and aspect system in English. They are past, present and Future tenses each has four aspects (simple, continuous, perfect, perfect continuous). At the Arabic side, there are three tenses of verbs which denote the sentence tense: the past, the present, and the imperative. The past expresses an event which happened before the moment of speaking. The present expresses an event which happens at the moment of speaking or after it. The imperative is considered in Arabic as a tense unlike English which considers it as a mood. The imperative expresses a request that will happen after the moment of speaking. However, Arabic consists of only two tenses which are not precise equivalents for various English tenses.

Table 2: The Arabic word order assignment for each English tense.

Serial no.	English word order	Tense	Arabic word order
1	S + V	Past simple	V + S
2	S + V	Past continuous	كان + S + V
3	S + V	Past perfect	قد + كان + S + V
4	S + V	Past perfect continuous	كان + S + V
5	S + V	Present simple	V + S
6	S + V	Present continuous	إن + S + V
7	S + V	Present perfect	قد + V + S
8	S + V	Present perfect continuous	يكون + S + V
9	S + V	Future simple	سوف + V + S
10	S + V	Future continuous	سوف + يكون + S + V
11	S + V	Future perfect	سوف + يكون + S + قد + V
12	S + V	Future perfect continuous	سوف + يكون + S + V

Control method: the selected the Arabic sentence word order structures SVO and VSO in combined with some fixed words is used to form a particular Arabic sentence word order for each tense in English. The aim of each modified Arabic sentence word order is to deliver the time of action specified by the English tense. The fixed words are:

– كان (kAn) and its different forms. It means be or been.
– سوف (swf) which means will or shall.
– إن (<n) which is a particle.

The following table shows the English different tenses and the corresponding modified

Arabic sentence word order structures. We demonstrate only the English sentences with SV order as an example.

3.3.3 Agreement divergence

Agreement happens when a word changes form depending on the other words to which it relates. It is an interconnection between words, especially marked by their inflections [15]. In English, the subject verb agreement exist only in few cases, mainly when using am, is, are, was, were and has to agree with the subject number. Also the verb agrees with subject when the verb is in the present tense for third-person, singular subjects. The adjective noun agreement does not exist in English. In general, Arabic verbs agree with their subjects in number and gender, and Adjectives agree with its nouns in definiteness, gender, number, and case. Some other agreements also exist in Arabic between the numbers and the countable nouns. There is a correlation between word order and verbal agreement in standard Arabic, i.e. full agreement in SVO order and partial agreement in VSO order.

14) They are beautiful boys.
هم أولاد جميلون (hm >wlAd jmylwn)
They(plural-masculine) boys(plural-masculine) beautiful(plural-masculine).

15) They are beautiful girls.
هن فتيات جميلات (hn ftyAt jmylat)
They (plural-feminine) girls (plural-feminine) beautiful (plural-feminine).

In both above examples, the subject pronouns agree with the subject complement gender and number, and the adjectives agree with their nouns in gender, number, case and definiteness. In the adjective noun and subject verb agreement relationship, there is a case when the adjective and verb are not totally agree with its noun/subject features and the agreement exclude the number and gender from the agreement features. This case is when the noun is not for a human. The adjective and the verb in this case are always on the form of a singular feminine; for example:

16) The stray dogs are running.
الكلاب الضاله تجري (AlklAb AlDAlh tjry)
The dogs (plural, masculine) the stray(singular, feminine) running (singular, feminine)

All gender and number affect the agreement feature which is very important in constructing the Arabic sentence. The error or problem caused by wrong assignment of the gender or number will result in incorrect or inappropriate words agreement and finally wrong translation output.

Control method: word agreement have occupied center stage in the translation process. Starting from the translation of the English sentence, the sentence is analysed by both parser and tagger to indicate the part of speech of each word. According to the output of the parser and tagger tool, the agreement relations are decided and the sentence is divided into chunks of related words. These chunks are mainly the subject, verb, direct object, indirect object and so on. The chunks are rearranged according to the Arabic selected word order. If the first chunk is not the subject then analyse the subject chunk and extract its main word features and then translate the chunks. The translation is done by translating the chunks one by one. Each chunk may contain more than one word. An agreement mostly exists between chunk words. When translating a chunk, information regarded any other related chunk is always present. The translation of a chunk starts from the most right word in the chunk where the main word/noun is located. Processing the main noun provides the necessary information needed to construct the chunk's remaining words and to fulfill the required agreements.

To control the irregular case of the adjective noun and subject verb agreement when the noun/subject is not for human, the Arabic noun attached features at the bilingual dictionary is set to singular feminine for all the dual and plural inhuman nouns such as (dogs → كلاب |tp #sing *feminine).

Most parts of the pseudo-code steps regarding the control of the agreement problem have been mentioned at previous subsections. Here we will list the pseudo-code of translating the verb chunk demonstrating the steps where the agreement relation is done.

```
If the chunk is the verb chunk
  Translate the verb
  Select the appropriate form according to the sentence tense, type and voice
  If the verb is following the subject Then
    Adjust the verb to agree with the subject in number and gender
    Else Adjust the verb to agree with the subject in gender
  Adjust the verb to match any other grammatical states
Add the translation to the sentence output translation
```

IV. IMPLEMENTATION

In order to minimize the errors rate during translation, the divergences and ambiguities control methods are applied throughout the translation process and in every module of our ANN and Rule based machine translation system. We used language analysing tools in addition to variables to save some data, and huge number of rules to tackle the problem of language divergences and ambiguity. For the source language, both the Parser and Tagger are used to analyse the English input. The Parser works out the grammatical structure of sentences, for instance, which groups of words go together (as "phrases") and which words are the subject or object of a verb. The Tagger is labeling a Part-of-Speech or other lexical class marker to each and every word in a sentence such as noun, verb, adjective, etc. The information

generated by both tools provide the backbone of the knowledge needed to translate the English text into Arabic. It is used by the rules in every decision making points, in solving of divergences and clarifying the ambiguities. For Arabic, the data at the ANN bilingual dictionary consist of Arabic words and an additional information about the word. The information is the word person, gender and number features. Word features are used to decide the chunk main word features, the selection of the right form of a word and fulfill the agreement. During translation, the order in which the English sentence chunks and the chunk words are processed for translation is a method to encounter the reordering and agreement divergences. The English sentence chunks are processed according to the Arabic word order while the chunk words are processed from the right most word which is the main word. Different types of variables are used to hold data which are needed at different stage of translation. The variables are used during the chunk, sentence and whole text translation. Throughout the translation, huge number of rules are used in every decision point.

V. RESULTS AND CONCLUSION

One important feature of machine translation is to maximize meaning, so that minimum effort and less time are required to understand the output. The user should not put forth too much effort to link together the different elements of the translation. In this paper, we have discussed some previous work determining the divergence and ambiguity of the English Arabic machine translation and its detection. The conventional type of divergences and ambiguities and their detection in English to Arabic MT have been presented. We have proposed implementation strategy that is used to detect and control divergences and ambiguities in our MT system. These strategies include Tagger and Parser tools, the data structure at the bilingual dictionary, and the use of rules and variables during translation.

Our MT system handles most of the types of divergences and ambiguities while translating from English to Arabic. In general, it is evident that it is not possible to deal with all kinds of divergences and ambiguities. However, most of the commonly found divergences and ambiguities have been discussed in this paper. Our MT system is ANN and rule based model. ANN is used as a bilingual dictionary. The MT system is able to handle conventional type of ambiguity such as lexical and syntactic ambiguity, and structural divergences. The lexical types of ambiguity are category, homograph, transfer, pronoun reference, and gender and number. The word order, tense and agreement are the three types of the structural divergences.

REFERENCES

[1] M. Baker, "In Other Words: A Coursebook on Translation," ed: Routledge (Londres), 1992.

[2] A. Pym, "Translation error analysis and the interface with language teaching," *Published in The Teaching of Translation. Ed. Cay Dollerup and Anne Loddegard, Amesterdam: John Benjamins,* pp. 279-288, 1992.

[3] A. Soudi, A. Farghaly, G. Neumann, and R. Zibib, *Challenges for Arabic Machine Translation* vol. 9: John Benjamins, 2012.

[4] N. Y. Habash, "Introduction to Arabic natural language processing," *Synthesis Lectures on Human Language Technologies,* vol. 3, pp. 1-187, 2010.

[5] M. Abu Shquier, "Word agreement and ordering in English-Arabic machine translation: a rule-based approach.," PhD PhD thesis, FTSM, University Kebangsaan Malaysia, Malysia, 2009.

[6] M. A. Attia, "Handling Arabic morphological and syntactic ambiguity within the LFG framework with a view to machine translation," the University of Manchester, 2008.

[7] J. Elming and N. Habash, "Syntactic reordering for English-Arabic phrase-based machine translation," in *Proceedings of the EACL 2009 Workshop on Computational Approaches to Semitic Languages*, 2009, pp. 69-77.

[8] M. Akeel and R. B. Mishra, "ANN and Rule Based Method for English to Arabic Machine Translation," *International Arab Journal of Information Technology*, vol. 11, July 2014.

[9] T. Buckwalter, "Buckwalter Arabic Morphologcal Analyzer," in *Linguistic Data Consortium. (LDC2002L49)*, 2002.

[10] D. Klein and C. D. Manning, "Accurate unlexicalized parsing," in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, 2003, pp. 423-430.

[11] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, 2003, pp. 173-180.

[12] W. J. Hutchins and H. L. Somers, *An introduction to machine translation*. London: Academic Press, 1992.

[13] W. J. Hutchins, "Milestones in the history of machine translation," in *Presentation at the conference "Current issues in theoretical*

and applied linguistics", Chelyabinsk, Russia, 2007.

[14] J. Elming, *Syntactic reordering in statistical machine translation*: PhD School in Economics and Business Administration, Copenhagen Business School, 2008.

[15] J. Algeo and T. Pyles, *The Origins and Development of the English Language* vol. 1: Cengage Learning, 2009.