

Investigating the Accuracy of Google Translate in Translating Urdu Linguistic Elements

Zafar Ullah¹,
Fakiha Iqbal Shah²

¹Department of English, National University of Modern Languages, Islamabad

²Department of English, National University of Modern Languages, Islamabad

Abstract

This study aims to investigate the accuracy of Google Translate regarding the translation of idioms and proverbs from and to English language. The shortcomings of Google Translate and the limitations of Neural Machine Translation have also been pointed out after in-depth analysis of carefully chosen examples from authentic lexicons. An experiment has been conducted by running Urdu and English idioms, proverbs and other linguistic expressions through Google Translate. The theoretical framework of Antoine Berman's twelve deforming tendencies has been employed to analyse these examples and Mixed Methods design has been applied for the evaluation in the study. The error analysis highlighted all the linguistic mistakes and errors committed by the software during Google machine translation. Major contribution of this study is the provision of the accurate alternatives for the wrong translation. This study plays a diagnostic role in identifying and classifying those errors. Furthermore, it contributes in improving the quality of Machine Translation from Urdu to English and English to Urdu. In the end, suggestions for the effective advancements in the software have also been provided.

Keywords: Google Neural Machine Translation (GNMT), Google Translate, Machine Translation (MT), Neural Machine Translation (NMT), Translation Errors.

Author's E-mail: fakiha.iqbal.96@gmail.com

Introduction

Translation is the process of communicating the meaning of a text from one language to another, hence, it requires great effort, precision and care so the exact meaning should be transferred without any omission, expansion or destruction. There are around 6,500 languages spoken and written all around the world. During contemporary era, the world has

transformed into a global village, translation has become one of the major means of international communication. Therefore, there is a dire need for automating the translation process and making it as accurate as possible. This research sheds light on the errors committed by translation software, specifically Google Translate, while translating certain linguistic elements from English to Urdu and Urdu to English.

Warren Weaver (1894-1978) is considered as the pioneer of machine translation due to his "Memorandum on Translation" written in 1947 (Hutchins, 1997). In the memorandum, he presented the idea of the field of machine translation. Practical efforts were initiated by Yehosha Bar-Hillel in 1951. Throughout the last half of the 20th century, MT progressed at a fast pace and by 1997, SYSTRAN was giving free translation services for short texts online. Consequently, in 2009, Google introduced Google Translate, its complimentary translation service. Google Translate was initially based on Statistical Machine Translation (SMT) which translated the text into English first, and then into the target language. It had poor accuracy and made several errors. Therefore, in 2016, Google switched to Neural Machine Translation (NMT) which involved deep learning and produced more accurate results. The end-to-end learning approach helped translate text from the source language directly to the target language after analysing millions of text samples. Due to this advancement, Google Translate now reportedly translates over 100 billion words a day, as of 2018.

Google has made it easier for its users to access this service by launching its application in 2011. It offers translational services in over 100 languages for free. It can also translate webpages, texts in pictures and handwritten text. In 2012, Google announced that around 200 million people use Google Translate every month for translating different kinds of

text. Now, over 200 million people use it daily. Therefore, improving its accuracy will benefit millions of users. Google also has its own Google Translate Community which encourages users to improve the accuracy of Google Translate by providing the most accurate translations of texts provided by Google itself. The process is not very effective as it is slow and allows users to only contribute for the words or phrases it asks for itself.

This study raises the following research question of where and how does Google Translate fail to translate commonly used Urdu proverbs, idioms and other linguistic structures from English to Urdu language and vice versa? To answer this question, certain proverbs and idioms have been translated to and from Urdu and, the results have been analysed qualitatively. These linguistic elements have been taken individually, without providing any context to the software, as the focus of this research is limited to common Urdu and English expressions. This has not only lead to the discovery of common errors committed by Google Translate but has also pointed out the shortcomings of the neural machine translation. As the field of machine translation is still young, finding such flaws will pave way for improvements and advancements. Such changes have the power to positively impact millions of people and make translation easier.

Literature Review

Arturo Trujillo (Trujillo, 1999) stated the history of Machine Translation in a succinct manner. The initial efforts to translate texts with the help of machines had been started in 1947 since the early use of computers for deciphering codes during World Wars. The first Russian-English prototype for Machine translation was developed in 1954 based on strategies proposed by Warren Weaver. This laid the foundation of machine translation of natural languages in the United States and around the world. During 1960s, machine translation faced a setback in government funding when it was proposed that Machine Translation was not cost effective. The funds were directed towards Artificial Intelligence and a few Machine Translation groups survived in the USA, with most of the research being done abroad. However, during the 1970s, SYSTRAN was advanced to the level where it was taken up to be used by US Air Force and the Commission of the European Union. By the 90s, the field had seen so much progress that it was being used around the world to translate academic and professional texts on personal computers. and in software, multimedia, personal computers and Shifting from SYSTRAN gradually to Statistical Machine Translation proved to be revolutionary for translational purposes. The Statistical Machine Translation (SMT) systems translate texts based on statistical models of bilingual corpora (Trujillo, 1999). In 2013, Aasim Ali, Arshad Hussain and Muhammad Kamran Malik studied the

Model for English-Urdu Statistical Machine Translation (Aasim Ali, 2013). Their study was based on the issues regarding corpus alignment as no parallel aligned data were available for these languages. They found that, due to Statistical Machine Translation's complete reliance on aligned data, words with low or none corresponding occurrences were not translated into the target language.

In 2014, a study was conducted regarding English to Urdu Statistical Machine Translation: Establishing a Baseline (Bushra Jawaid, 2014). The research compared baseline Phrase Based Machine Translation (PBMT) with Hierarchical Machine Translation and found the latter to be more accurate for Urdu-English translations.

In 2016, Google switched from Statistical Machine Translation (SMT) to Neural Machine Translation (NMT) (Yonghui Wu, 2016). It claimed that due to its end-to-end learning approach, it was more accurate than SMT. They called it Google Neural Machine Translation (GNMT) and concluded that due to its accuracy and speed, GNMT could translate huge amounts of data in high quality.

Several researches regarding Neural Machine Translation followed and worked on finding the limitations of NMT. A research, in 2017, regarding Fully Character-Level Neural Machine Translation without Explicit Segmentation concluded that using fully character-level NMT helped translate words without explicit

segmentation and the model learnt these concepts from the act of translation itself (Jason Lee, 2017). This model was suitable for languages with rich morphology and vocabulary such as Urdu.

Very little work had been done regarding Urdu-English machine translation with the help of Google Translate. A couple of notable researches included a study of Urdu to English Machine Translation using Bilingual Evaluation Understudy (Asad Abdul Malik, 2013). It focused on the accuracy of SMT in comparison to EBMT and concluded that EBMT proved to be more accurate when translating Urdu to English.

Another important contribution was made in 2014 when Sharmin Muzaffar, Pitambar Behera brought out the issues pertaining to the plural markers of the Urdu verbs based on data generated through Bing and Google translators. With a view to capture the verbal inflections, the data for experimentation purpose was collected from natural language for simple forms of all tenses (Muzaffar & Behera, 2014). Typical Urdu last words for 12 tenses (TA HE, TE HE, TA HEY for Simple Present Tense) must be added for correct translation or detecting the right tense in Urdu.

All the aforementioned studies contributed in the field of Machine Translation but none of them dealt with the Urdu-English translation by NMT or GNMT. Moreover, this study focused on the translation of selected

idioms, proverbs and other commonly used linguistic expressions in both English and Urdu languages. Therefore, this study is unique and original as compared to all the previous studies.

Research Methodology

The purpose of this research was to test the accuracy and realism in the translation by Google Translate's neural machine system. For this purpose, we used an inductive approach and induced results from the selected examples. Moreover, mixed methods were used for this research as it allowed the use of both qualitative and quantitative data. Mixed methods also made the research more accurate as the use of mixed methodologies and cross-checking is possible. The main objective of this research was to find the shortcomings of Google Translate so its creators can improve its Urdu-English translation according to authentic lexicons like Oxford Urdu English Dictionary and Farhang-e-Aasifiya.

The data used in this exploratory research had been taken from different sources for cross-validation with the help of simple random sampling. The sources of the data are authentic publications by the Oxford University Press. The total corpus consists of sixty items from which ten English idioms were taken from Seidl & McMordie (1988) and ten Urdu idioms were taken from Phillott (1912). For English proverbs, Speake (2008) was taken as source and ten proverbs were taken from here.

Furthermore, ten Urdu proverbs were taken from 'Famous Urdu Proverbs Translated into English' (2014). Moreover, other lexical elements including grammatical gender and relations were also translated. The first section focuses on the results generated by translation of English text to Urdu and conversely, the second section focuses on Outcomes obtained by translating Urdu text to English.

The translation of these idioms, proverbs and other lexical expressions were then analysed on the basis of the Theoretical Framework of Antonio Berman's twelve deforming tendencies (Berman, 2000). These tendencies, according to Berman, were innate in translation and were, therefore, unavoidable. These twelve tendencies include rationalisation, clarification,

expansion, ennoblement or popularisation, qualitative impoverishment, quantitative impoverishment, the destruction of rhythms, the destruction of underlying networks of signification, the destruction of linguistic patterning, the destruction of vernacular network or their exoticisation, the destruction of expressions and idioms, the effacement of the superimposition of languages. The study finds out which tendencies are more dominant than others in translation by Google Translate.

Results and Discussion

Following are ten English idioms that have been taken from Seidl & McMordie (J. Seidl, 1988) and translated through the Google Neural Machine Translation software:

Table English Idioms

Sr. No.	Source Language	Target Language
1	Bad blood	گندا خون
2	To be in someone's bad books	کسی کی بری کتابوں میں شامل ہونا۔
3	To be hard on someone	کسی پر سختی کرنا۔
4	Through thick and thin	موٹی اور پتلی کے ذریعے۔
5	A thin excuse	ایک باریک عذر۔
6	First things first	ضروری کام پہلے
7	Small world	چھوٹی دنیا
8	Once and for all	ایک بار اور سب کے لئے
9	There and then	وہاں اور پھر۔
10	A breath of fresh air	تازہ ہوا کا ایک سانس۔

In the very first example, Google Translate has translated the idiom literally, word to word. Moreover, the translation 'bad blood' has a completely different connotation in Urdu than the meaning of 'bad blood' and is a separate term. Therefore, the original meaning of the idiom is lost in translation. Similarly, in the second example, the translator has carried out literal translation of the idiom and hence, loses its original meaning. In the third example, it carries out correct translation of the idiom. In the fourth example, it again carries out literal translation and uses feminine terms for describing 'thick and thin'. This choice of grammatical gender exhibits destruction of the vernacular network. The same problem is with the fifth example, although, here it uses the word 'thick' for 'thin' which is totally inaccurate. This adjective cannot be used with the word 'thin'. Therefore, it not only fails to convey the correct meaning of the idiom, but it also ends up constructing an expression without any proper meaning. Similarly, in the eighth and ninth example, it again carries out literal translation destroying

the meaning of the idiom completely, however, it does accurate translation of the sixth, seventh and tenth example.

The main problem with GNMT is its inability to convey the connotative meaning of the idioms. It keeps on producing results with denotative meanings, by carrying out word to word translations. Moreover, it fails to construct any meaningful sentence in the target language as it ignores the semantic properties of the target language. Therefore, it seems to fail to switch to the semantic rules of the target language and is instead projecting the semantics of the source language on the target language. For example, meaning of 'through thick and thin' (Seidl & McMordie, 1988) is 'through good times and difficult times, in all conditions' which is not in any way conveyed by the translation 'موٹے اور پتلے کے ذریعے', thus making the translation wrong and useless.

Ten English proverbs from (J. Seidl, 1988) (Speake, 2008) have been translated through BMT. Following are the generated results.

Table 2 English Proverbs

Sr. No.	Source Language	Target Language
11.	Marriages are made in heaven.	شادیوں جنت میں ہوتی ہیں۔
12.	Laughter is the best medicine.	ہنسی بہترین دوا ہے۔
13.	Jack of all trades and master of none.	تمام تجارت کا جیک اور کسی کا ماسٹر نہیں۔
14.	As you sow, so you reap.	جیسا کہ آپ بونے ہیں، لہذا آپ کو کاٹنا پڑتا ہے۔
15.	Prevention is better than cure.	پرہیز علاج سے بہتر ہے۔
16.	United we stand, divided we fall.	متحد ہوں گے تو قائم رہیں گے، تقسیم ہوں گے تو شکست خوردہ ہوں گے۔
17.	Bad things come in threes.	بری چیزیں آتی ہیں۔
18.	A bad workman blames his tools.	ایک خراب کاریگر اپنے اوزار کو مورد الزام ٹھہراتا ہے۔
19.	Charity begins at home.	خیرات گھر سے شروع ہوتی ہے
20.	A barking dog never bites.	بھونکنے والا کتا کبھی نہیں کاٹتا۔

The eleventh proverb has been translated verbatim from English to Urdu although the correct translation is جوڑے آسمانوں پر بنتے ہیں۔ Moreover, the translation of the word 'marriages' is also wrong, showing how the software has trouble translating plural nouns. The twelfth proverb has also been translated word-to-word instead of translating it into corresponding Urdu proverb ہنسی علاج غم ہے۔ The translation of thirteenth proverb has several errors in it. It has transliterated the words "jack" and "master" into Urdu. Moreover, the tool has misinterpreted the whole meaning and has given the literal translation of the proverb. Similarly, in the fourteenth proverb, a

very common proverb has been translated as it is rather than in its corresponding Urdu proverb جو بوگے۔ وہ کاٹے۔ The translation of the fifteenth proverb is correct. In the sixteenth proverb, the software has not only literally translated the whole proverb to Urdu but has also changed the tense and expression. In the next example, the translation is literal as well as incomplete. The expression 'in threes' has not been translated at all and only the first part of the proverb has been translated. The last three examples also include verbatim translation instead of translation to corresponding Urdu proverbs.

*Investigating the Accuracy of Google
Translate in Translating Urdu Linguistic Elements*

For the next set of examples, ten Urdu idioms were taken from Phillott (1912) (Phillott, 1912) and translated

through Google Translate. The following table contains the obtained results.

Table 3 Urdu Idioms

Sr. No.	Source Language	Target Language
21.	آسمان سے باتیں کرنا	Talking to the sky
22.	آسمان کے تارے توڑنا	Break the stars of the sky
23.	بھیگی بلی	Soaked pussy
24.	پانی پانی کرنا	Water to water
25.	پتھر کا دل	Heart of stone
26.	ٹیڑھی کھیر	Crooked pudding
27.	خاطر میں نہ لانا	Do not bring in the sake of
28.	خون سفید ہونا	Blood to be white
29.	دن پھرنا	To roam the day
30.	ذندہ درگور ہونا	To be mindful

In this table, examples 21, 22 and 23 have all been literally translated in English although their corresponding English idioms do exist. The translations do not make any sense and are therefore, semantically wrong. In the 24th example, the English translation is not only incorrect, but it has also changed the gerund کرنا to the preposition 'to'. The examples 25 and 26 are again cases of poor verbatim copy and a lack of rationalization has been observed in the latter. However, in the 27th example, the literal translation is also wrong. The negation verb نہ لانا has been changed into an imperative 'do not' showing the inaccuracy of the denotative meaning. In example 28, the translation makes no sense and lacks

clarity of meaning. In example 29, again a deficiency of meaning can be observed, resulting in an incorrect expression. In the 30th example, the translation is completely inappropriate and fails to provide the connotative meaning.

Here, ten Urdu proverbs have been taken from 'Famous Urdu Proverbs Translated into English', (2014) (Best Right Way, 2014) and were translated by Google Translate:

Table 4 Urdu Proverbs

Sr. No.	Source Language	Target Language
31.	کھسیانی بلی کھمبا نوچے	Khasani cat poles dancing
32.	چھوٹا منہ بڑی بات	Small mouth big thing
33.	دور کے ڈھول سہانے	the grass is always greener from the other side
34.	دھوبی کا کتا نہ گھر کا نہ گھاٹ کا	Dhobi's dog neither the house nor the pier
35.	جیسی کرنی ویسی بھرنی	what goes around comes around
36.	جتنی چادر ہو اتنا پیر پھیلاو	Spread the legs as wide as possible
37.	اب پچھتائے کیا ہوت ، جب چڑیاں چگ گئیں کھیت	Now what would be the regrets, when the sparrows were consumed
38.	جو گرجتے ہیں وہ برستے نہیں	Those who roar do not fall
39.	لالچ بری بلا ہے	Greed is evil
40.	ڈوبتے کو تنکے کا سہارا	Resort to straw to drown

In the very 41st example, the word کھسیانی has been transliterated whereas the word نوچے has been wrongly translated as 'dancing'. In the 42nd example, the proverb has been translated verbatim and the translation of بات has been mistranslated as 'thing'. The 43rd example has been translated into the English proverb but lacks accuracy as the correct expression is 'grass is always greener on the other side', not 'from the other side'. The next proverb has also been incorrectly translated. The word دھوبی has been transliterated although possession has been shown with the help of '-s', while the rest of the proverb has been translated word-to-word. The fifth proverb has also been translated incorrectly. Although, this translation also corresponds to this proverb, still

the accurate translation is 'as you sow so shall you reap'. The next translation is entirely wrong as the English expression does make perfect sense but does not correspond to the Urdu proverb at all, rather holds a meaning quite opposite to the actual meaning which is to undertake only what one has the ability to do. The next translation is also an example of poor verbatim copy. The software has also changed the verb 'regret' to noun 'regrets' and the subject 'sparrows' has been converted into receiver of the action. The example shows poor sense of syntactic pattern of the software. The 48th proverb has also been translated word to word, and that too, incorrectly. The correct translation is 'barking dogs seldom bite'. Moreover, the word برستے has wrongly translated as 'fall'. The ninth proverb is again translated verbatim although the

correct translation should be 'avarice is the root of all evils. The last example also has a distorted verbatim translation whereas the correct translation is 'drowning man catches at straw'.

The following table contains a list of sentences mentioning different relations in Urdu and their translations by Google Translate.

Table 5 Relations

Sr. No.	Source Language	Target Language
41.	وہ میرے نانا ہیں۔	He is my grandfather.
42.	وہ میرا بہنوئی ہے۔	He is my sister.
43.	یہ میرے خالو ہیں۔	This is my uncle.
44.	وہ میرے چچا کا بیٹا ہے۔	He is my uncle's son.
45.	وہ میری پھوپھی کا بیٹا ہے۔	She is my aunt's son.
46.	وہ میرا منہ بولا بھائی ہے۔	He is my brother.
47.	وہ میرا نندوئی ہے۔	That's my ninety.
48.	میں اس کا لے پالک بیٹا ہوں۔	I am his adopted son.
49.	وہ میرا خالہ زاد ہے۔	He is my cousin.
50.	وہ اس کے نانا کی بھانجی ہے۔	She is her grandfather's niece.

The first sentence has been translated as 'he is my grandfather' which is ambiguous as **نانا** is the maternal grandfather and the translation fails to mention that. In the 52nd example, the relation of **بہنوئی** has been mentioned which refers to 'brother-in-law' or more specifically, 'sister's husband' but the translation is totally incorrect, and the software has used 'he', a masculine pronoun to refer to 'sister', a feminine noun. The 53rd sentence also has the problem of ambiguity as the word **خالو** specifically refers to 'maternal aunt's husband' but it has been simply stated as 'uncle' leaving confusion and giving incomplete meaning. The 54th example again lacks complete information as it simply

translates **چچا کا بیٹا** to 'uncle's son', failing to state if the uncle is paternal or maternal. The 55th sentence has been mistranslated and the software has again juxtaposed a feminine pronoun 'she' with a masculine noun 'son'. Moreover, the translation of **پھوپھی** has been simply done as 'aunt' which fails to communicate whether it is maternal or paternal aunt. The 56th translation is an example of incomplete translation as the relation **منہ بولا بھائی** has been simply translated as the word 'brother' which is incorrect. The 57th example is of the relation of **نندوئی** who is 'sister-in-law's husband' or 'husband of one's husband's sister'. Google Translate fails to understand this word and mistranslates it to ninety, rendering the whole

sentence meaningless. The 58th sentence has been translated correctly. The 59th sentence is again translated incompletely as the word *خالہ زاد* has simply been translated to cousin without mentioning which cousin. The last translation has a surprisingly odd mistake as Google Translate

successfully translates all relations correctly but fails to translate the pronoun correctly. The pronoun *اس* is gender neutral in Urdu and can be used for both masculine and feminine pronouns but Translate restricts the translation to only 'her', classifying it as a feminine pronoun solely.

Table 6 Linguistic Elements

Sr. No.	Source Language	Target Language
51.	اس کا نام میرے نام پر رکھا گیا۔	It was named after me.
52.	وہ دانت نکال رہا تھا۔	He was pulling his teeth.
53.	اس نے پھول توڑا۔	He broke the flower.
54.	ستارے آسمان پر ہیں۔	The stars are in the sky.
55.	ڈاکٹر مریض کی نبض دیکھتا ہے۔	The doctor watches the patient's pulse.
56.	کبوتر کے پر کاٹ دو۔	Cut it on the pigeon.
57.	وہ جاتے جاتے رہ گیا۔	He kept on going.
58.	اس نے اس کو پانی پلایا۔	He gave her water.
59.	ناک صاف کرو۔	Clean the nose.
60.	نالے دانی لاؤ۔	Bring the umbilical cord.

The above table consists of examples regarding commonly used linguistic elements of Urdu language. Some of these include multiple third-person pronouns and others employ collocations. In the first example, the software translates “اس” to “it” although there is no indicator of whether “اس” refers to an animate or an inanimate object. In the 62nd example, the Urdu sentence correctly translates to “he was grinning” but Google translate has literally translated the expression *دانت نکال رہا تھا* to “pulling his teeth”. In the 63rd example, the word *توڑا* here means “plucked” but the

software wrongly translates it to “broke” regardless of the fact that the word has been used in another context. The 64th sentence has been translated correctly. The fifth example is again verbatim translation and fails to convey the correct meaning. The correct translation should be “The doctor checks the patient's pulse” but the software fails to take into account the collocations regarding the word *دیکھتا*. In the 66th example, the translation is incorrect as the software wrongly translates *پر* to “on” rather than “wings”. Google Translate fails yet

again to analyse the context of the word. The 67th example shows that the software has wrongly translated the Urdu sentence, resulting in complete loss of intended meaning. in the 68th example, the software translates the “اس” in the subject to “he” and “اس” in the object to “her” although there is no indicator of grammatical gender in the sentence. The 69th example again shows how Google Translate fails to understand the context of the word صاف and translates it to “clean” rather than “wipe”. In the last example, the word نالے دانی refers to a tool used for putting elastic in the shalwar. Due to cultural differences, there is no corresponding word for نالے دانی in English. Hence, Google Translate wrongly translates it to “umbilical cord”.

The major findings of this study show the different types of errors committed by the software. The software does not have several corresponding Urdu words in English and transliterates them, such as کھسیانی and دھوبی. Moreover, certain English words have also been transliterated to Urdu such as 'Jack' and 'Master'. The software also fails to maintain coherence among nouns and pronouns and lacks vocabulary regarding relations.

Major Findings

Major findings in this research are:

Semantic Errors: The translations include semantic errors which refer to the lack of meaning in translated

expressions and fail to produce meaningful sentences at times. Strictly considering the data above, the frequency of semantic errors in translation was 35% such that the whole expression was meaningless. Moreover, the frequency of partial errors where some sense of meaning is retained in translation and the error was only due to one-word errors was 8.33%. For instance, the expression “through thick and thin” was translated to موٹی اور پتلی کے ذریعے which does not have any meaning in Urdu language and چھوٹا منہ بڑی بات was translated to “small mouth big thing” where some meaning is conveyed but lack of use of appropriate words made it ambiguous.

Transliteration: Transliteration is the process of rewriting a word of a language in the alphabet of another language. The frequency of this error was 5% as there were only a few words in the above examples which Google Translate failed to translate. Whenever the software failed to recognize an Urdu word, it transliterated it from Urdu to English. For example, the words کھسیانی and دھوبی were transliterated to “khasani” and “dhobi” respectively.

Verbatim Translation: A verbatim translation is the word-to-word copy or translation of a word or

expression from one language to another. Google Translate carried out verbatim translation of majority of the idioms and proverbs, consequently losing the real meaning. the frequency of this error was 62.5% which is very high. For example, آسمان سے باتیں کرنا was translated verbatim as “talking to the sky” and “United we stand, divided we fall” as متحد ہوں گے تو قائم رہیں گے، تقسیم ہوں گے تو شکست خوردہ ہوں گے.

Grammatical Gender Problems: Urdu language is based on the use of grammatical gender due to which specific pronouns are used with certain nouns, based on their grammatical gender. Google Translate fails to account for this rule and incorrectly translates expressions. There was a 10% frequency of the error in which the software got the gender completely incorrect while 15% of the time it chose gender neutral terms in translation, making the results ambiguous. Some examples include وہ میری پھوپھی کا بیٹا ہے becomes “She is my aunt's son” and وہ میرا بہنوئی ہے is translated to “He is my sister”.

Lexical Limitations: English language lacks the equivalents of certain Urdu words due to socio-cultural differences. The software either fails to find the appropriate translation of certain words or simply does not have the required data. This

can be seen through the examples of words like نالے دانی, نندوئ and بہنوئ for which the software fails to provide appropriate meanings and instead comes up with wrong translations which are “umbilical cord”, “sister” and “ninety”, respectively. The frequency of this error was recorded to be 15%. There were also some partial errors where the software recognized the words from source language correctly but not completely such as “marriages” was translated to شادیوں and not شادیاں. The frequency of such errors, on the basis of the above data, was recorded to be 11.67%. Other times, it simply transliterates those words, for example, “master” remains ماسٹر, and دھوبی remains “dhobi”.

Misinterpretation by the Tool: Misinterpretation refers to error in understanding the whole expression. It sometimes results in devising a meaning quite opposite to the real meaning. This error is evident in the results as the software translates جتنی چادر ہو اتنا پیر پھیلاو to “spread the legs as wide as possible”.

Missing Syntactic Pattern: Google Translate failed several times to produce proper sentences during translation, thus resulted in poor syntactic patterns. For instance, ٹوہتے کو کا سہارا was translated to

“resort to straw to drown” which has no proper structure or meaning. Similarly, پانی پانی کرنا is translated to “water to water” which is an incomplete and meaningless expression.

English to Urdu and Urdu to English one-word translation was correct to some extent but it is not beneficial as one-word translation is least required by the users.

Conclusions

The aforementioned findings prove that Google Translate fails to accurately translate idioms, proverbs and many other commonly used linguistic elements correctly from Urdu to English and vice versa. Google Translation needs to be improved to acknowledge lexico-semantic elements, context and cultural connotations of words to translate idioms, proverbs and other linguistic expressions accurately. Besides, this analytical study would be beneficial in terms of improvement in efficiency of Google Translate as this study highlights translation errors which are responsible for deterioration of machine translation process. The application of machine translation in the translation of webpages, Skype conversations, Instagram captions, YouTube subtitles, Artificial Intelligence, conversations with robots in dialog system, reviews and comments on commercial websites has made it extremely necessary to be accurate. In addition, overall impact of Urdu language would be improved in

international community especially Pakistani Urdu variety will get more impetus and exposure. Furthermore, beneficiaries of this study would be translators, translation agencies and students of literature, linguistics and all other subjects. It will also be of great service to the learners of Urdu language. Further work needs to be done in order to improve the transfer of connotative meaning during the process of Neural Machine Translation of English-Urdu language pair, for the reason that “All translation is, and must be, the restitution of meaning” (Berman 2000, p. 297).

References

- AbdulMalik, A., & Habib, A. (2013). Urdu to English machine translation using bilingual evaluation understudy. *International Journal of Computer Applications*, 82(7), 5-12. doi:10.5120/14126-1040
- Ali, A., Hussain, A., & Malik, M. K. (n.d.). Model for English-Urdu Statistical Machine Translation. *World Applied Sciences Journal*, 24(10), 1362-1367. doi:10.5829/idosi.wasj.2013.24.10.760
- Famous Urdu proverbs translated into English. (2014, April 16). Retrieved November 10, 2019, from <https://www.bestrightway.com/2014/04/famous-urdu-proverbs-translated-into.html>

- Hutchins, J. (1997). From first conception to first demonstration: the nascent years of machine translation, 1947-1954 A chronology. *Machine Translation*, 12(3), 195-252. Retrieved from <https://doi.org/10.1023/A:1007969630568>
- Jawaid, B., Kamran, A., & Bojar, O. (2014). English to Urdu statistical machine translation: Establishing a baseline. *Proceedings of the Fifth Workshop on South and Southeast Asian Natural Language Processing*, 37-42. doi:10.3115/v1/w14-5505
- Lee, J., Cho, K., & Hofmann, T. (2017). Fully character-level neural machine translation without explicit segmentation. *Transactions of the Association for Computational Linguistics*, 5, 365-378. doi:10.1162/tacl_a_00067
- Muzaffar, S., & Behera, P. (2018). A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems. *Language in India*, 18, 154-163. Retrieved from <https://www.researchgate.net/define/http%3A%2F%2Fwww.languageinindia.com%2F>
- Phillott, D. C., & Dihlavī, S. A. (1912). *Khazīna-E Muḥāwarāt: Or, Urdu idioms*. Calcutta: Baptist Mission Press.
- Seidl, J., & McMordie, W. (1988). *English idioms*. New York: Oxford University Press.
- Speake, J., & Simpson, J. (2008). *A dictionary of proverbs* (6th ed.). Oxford University Press.
- Trujillo, A. (1999). *Translation engines: Techniques for machine translation* (1st ed.). London: Springer-Verlag London.
- Venuti, L., Baker, M., & Berman, A. (2000). TRANSLATION AND THE TRIALS OF THE FOREIGN. In *The translation studies reader*. London: Psychology Press.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... Dean, J. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.

