MEGA INSTITUTE FOR ADVANCE RESEARCH AND DEVELOPMENT

Journal of Arts and Linguistics Studies
ISSN Print 2959-0957 ISSN Online 2959-0965
Volume 4 Issue 7 2026
Pages 431-460
Open Access

Comparative Analysis of Digital Tools for Urdu-English Literary Text Translation: A BLEU Metric Evaluation

Publication Details
Paper Received:
May 10, 2026
Paper Accepted:
July 22, 2026
Paper Published:
July 24, 2026

Khadija Mushtaq
MS English Linguistics, International Islamic University, Islamabad.
E-mail: khadija4mushtaq@gmail.com
Dr. Rafia Bilal
Assistant Professor, Department of English, International Islamic University, Islamabad. Email: rafia.bilal@iiu.edu.pk

Article DOI: https://doi.org/10.71281/jals.v4i7.722

Abstract
This research conducted a comparative analysis of three digital translation tools, Google Translate, Lingvanex, and MateCat, evaluating their performance in translating Saadat Hasan Manto's Urdu short story "Khol Do" into English. The study utilized the BLEU (Bilingual Evaluation Understudy) metric to quantify the accuracy of each translation in relation to a human-translated reference translation by M. Umar Memon. Additionally, the research employed Molina and Albir's translation techniques to assess the quality of the translations. The goal was to identify the tool that consistently produced translations closest to the intended meaning and style of the source text or human translation. The findings indicated that Google Translate is the most successful tool in terms of both quantitative measures (having the highest BLEU scores) and qualitative aspects (having higher lexical fidelity and structural accuracy). The study highlighted the strengths and limitations of each digital translation tool in capturing the nuances of Urdu literature. The results suggested that while machine translation tools are improving, they still require human oversight to ensure accuracy and cultural relevance. The research contributed to the ongoing evaluation of machine translation tools and their applicability in translating literary texts, particularly from languages like Urdu, which posed unique challenges due to their linguistic and cultural specificities.

Keywords: Machine translation, Literary Translation, Molina and Albir Translation Techniques, Lingvanex, MateCat, Google Translate.

Published by MIARD. Copyright: © the author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

With the rising usage of machine translation (MT), multilingual communication has been revolutionized through instant and easy translation of text between languages. The problem of translating literary works is, however, a complex one since literary translations are full of cultural allusions, idioms, styles, and context that cannot be captured easily by automated tools. This problem is especially apparent in the case of Urdu to English literary translation.

Thus, the evaluation of machine translation systems has become one of the key areas of research. Among various automatic evaluation measures, the BLEU measure is popularly used to evaluate the quality of translations produced by machines by comparing them with human translation. Besides the quantitative measures, the qualitative assessment of the techniques of translation helps to understand better how the machine translation systems handle linguistic and cultural issues. The framework proposed by Molina & Albir (2002) can be useful for this purpose.

While earlier studies have focused on determining the effectiveness of machine translation either using automatic measures or analyzing translation methods independently, there have been very few comparative studies that have incorporated both of these factors for literary translation from Urdu into English. Thus, the current study compares the performance of three different tools, Google Translate, Lingvanex, and MateCat, in the translation of Saadat Hasan Manto's 'Khol Do' to English, with reference to the translation made by Umar Memon.

1.1. Statement of the Problem

Machine translation (MT) has been advancing at a rapid pace and its use for Urdu-English translation is increasing. However, its effectiveness for literary translation remains questionable, especially in the translation of cultural references, idiomatic expressions, emotional nuances and contextual meanings. This restriction

becomes very important in case of Urdu literature which is rich in culture and where authenticity of source text is very important. Digital translation tools are being used increasingly, but little research has comparatively been done on the performance of these tools in the translation of Urdu–English literature. Therefore, this study addresses this gap by comparing Google Translate, Lingvanex, and MateCat in the translation of Saadat Hasan Manto's Khol Do, using BLEU unigram precision and Molina and Albir's translation techniques to assess their accuracy and translation strategies against a human reference translation. The objective of the study is to find out the strengths and limitations of these tools and their applicability to Urdu–English literary translation.

1.2. Research Objectives

- To identify the techniques of translation employed by each of the tools.
- To evaluate and compare the accuracy of the translations generated by these tools using the BLEU metric.

Research Questions

1) Which translation techniques are employed by each of the tools?
2) Which of the three digital tools, Google Translate, Lingvanex, or Matecat, produces the most accurate translation as measured by the BLEU metric?

2. Literature Review

Machine translation (MT) has progressed from rule-based and statistical methods to Neural Machine Translation (NMT) that uses deep learning to enhance translation fluency and accuracy (Hutchins, 2005; Zong, 2018). However, literary translation is still a difficult task because of figurative language, cultural references, idiomatic expressions, stylistic features, and emotional nuances in literary texts which have to be interpreted in context. MT systems may generate lexical, syntactic, and cultural errors that impact both the meaning and the artistic

style of literary works (Besacier, 2014). Literary translation must be more than linguistic equivalence, it must preserve the esthetic and cultural dimensions of the source text. Baker (1992) states that idioms and culture bound expressions cannot be literally translated all the time since their meanings are not restricted to their individual words. Similarly, Nida's (1964) differentiation of formal and dynamic equivalence emphasizes the necessity of preserving the features of the source text and of creating an effect on the target reader.

Molina and Albir's (2002) framework of translation procedures has been frequently used in translation studies to investigate translation beyond mere lexical equivalency. Literal translation, borrowing, modulation, amplification, reduction, adaptation, transposition, and established equivalence are among the eighteen techniques identified by their framework. These methods have been used in earlier research to evaluate literary translations and contrast the results of human and machine translation (Yuliasri, 2016; Nduru, 2017; Sipayung et al., 2021). According to their findings, machine translation typically favors lexical correspondence and literal translation, but culturally and contextually appropriate translation frequently calls for more adaptable techniques.

Both human evaluation and automated measures are frequently used to assess translation quality. Human review is frequently laborious and subjective, even though it can evaluate fluency, sufficiency, and cultural appropriateness. Papineni et al. (2002) created the Bilingual Evaluation Understudy (BLEU) metric, which computes n-gram overlap between machine-generated and reference translations to provide an effective quantitative assessment. BLEU is nevertheless effective for measuring lexical similarity and can be improved through qualitative examination of translation procedures, despite its inability to fully represent stylistic, semantic, or cultural equivalency.

These difficulties are especially pertinent to translating from Urdu to English. The syntax, morphology, and word order of Urdu are very different from those of English, and it is regarded as a comparatively low-resource language. MT performance may be impacted by linguistic disparities and a lack of multilingual resources (Rauf et al., 2019; Raza et al., 2021). Literary writings are still challenging due to their culturally specific phrases, idioms, and emotional depth, even though NMT advancements have improved the quality of Urdu–English translations (Shah et al., 2020). Although digital translation systems like Google Translate, MateCat, and Lingvanex are becoming more widely available, there is still a dearth of comparative studies analyzing their effectiveness in Urdu-English literary translation.

Saadat Hasan Manto's Khol Do provides an appropriate text for examining these challenges because of its culturally embedded language, idiomatic expressions, and emotionally complex narrative. Although the story has drawn a lot of literary and translation-related attention, little research has looked into how modern MT technologies replicate its grammatical and cultural subtleties in English. Few studies have combined quantitative measurement with analysis of translation procedures; most existing research has concentrated on either automatic evaluation or qualitative assessment. In order to close this gap, the current study compares the translation of Khol Do from Urdu into English using Google Translate, MateCat, and Lingvanex.

3. Methodology

This study adopted a mixed-methods research approach, combining quantitative and qualitative methods to evaluate the performance of selected Urdu–English machine translation tools. Quantitative analysis was carried out using the BLEU metric (unigram precision) to measure lexical accuracy, while qualitative analysis was conducted through the application of Molina and Albir's (2002) translation

techniques to examine how linguistic and cultural elements were rendered in the target texts. This combined approach enabled a more comprehensive assessment of machine translation quality in short story "*Khol Do*" by *Saadat Hassan Manto*". The study was grounded in Molina and Albir's (2002) classification of translation techniques, which comprises eighteen categories ranging from direct techniques, such as borrowing, calque, and literal translation, to indirect techniques, including modulation, adaptation, and compensation. This framework provided a structured basis for identifying and comparing the translation strategies employed by Google Translate, Lingvanex, and MateCat in translating the selected literary text. Umar Memon's published English translation was used as the reference translation.

The complete dataset consisted of 24 sentences, and all sentences were analyzed quantitatively using BLEU and qualitatively using Molina and Albir's translation techniques. Due to the extensive length of the story, five sentences were selected as an illustrative excerpt for detailed sentence- level analysis. This analysis demonstrates how the three machine translation tools Google Translate, Lingvanex, and MateCat were evaluated using the BLEU metric and Molina and Albir's translation techniques. The excerpt therefore, serves as a representative demonstration of the analytical procedure, while the same procedure was applied to the remaining sentences in the complete dataset.

4. Data Analysis

This analysis was conducted on the Urdu short story "Khol Do" authored by Saadat Hasan Manto, along with its reference English translation done by Umar Memon. Three digital machine translation systems, namely, Google Translate, Lingvanex, and MateCat, were tested through this story. The short story was divided into 24 sentence-level units of analysis. The quantitative assessment was done through unigram precision feature of Bilingual Evaluation Understudy (BLEU) by matching the machine-translated sentence with the reference

translation. The process of computing BLEU involves tokenization of sentence, unigram precision, brevity penalty, and finally BLEU score. In addition to that, the qualitative analysis has been done by utilizing Molina and Albir's translation techniques framework for analyzing the translation techniques used by the three machine translation systems. Both the methods of analysis help in giving a detailed overview about the quality and strategies adopted by the chosen machine translation systems. The BLEU metric formula used for analysis is given below.

BLEU=BP× Precision

BP (Brevity Penalty) = If MT length > Reference length → BP = 1

If MT length ≤ Reference length → BP=e1−r/c

Where **r = reference length**, **c = candidate (MT) length**

Precision = Unigram matches ÷ MT token count

Analysis of Sentence #1

SOURCE TEXT	**REFERENCE TRANSLATION (HUMAN TRANSLATION)**	**GOOGLE TRANSLATE (TARGET TEXT 1)**	**LINGVANEX (TARGET TEXT 2)**	**MATECAT (TARGET TEXT 3)**
امر تسر سے اسپیشل ٹرین دوپہر دو بجے کو چلی آٹھ گھنٹوں کے بعد مغل پورہ پہنچی، راستے میں کئی آدمی مارے گئے	The special train left Amritsar at two in the afternoon, taking eight hours to reach Mughalpura. Quite a few passengers were killed along the way.	The special train from Amar Tasar left at 2 pm and reached Mughalpura after eight hours, many people were killed on the way.	The special train from Amar Tasar ran at two in the afternoon and reached Mughalpura after eight hours, many men were killed on the way.	A special train from Amar Tasar reached Mughalpura at 2 pm after eight hours, many people were killed on the way.

Quantitative Analysis The text has been analysed keeping in consideration Bleu metric analysis and its parameters.

a) Reference Translation The reference translation is presented in tokenized form and functions as the standard against which the outputs of the machine translation tools are evaluated. Total tokens of RT are 25.

["The", "special", "train", "left", "Amritsar", "at", "two", "in", "the", "afternoon", "taking", "eight", "hours", "to", "reach", "Mughalpura", "Quite", "a", "few", "passengers", "were", "killed", "along", "the", "way"] = 25

b) Google Translate Google Translate output is provided in tokenized form in order to assess its degree of correspondence with the reference translation. Total tokens of GT are 22.

["The", "special", "train", "from", "Amar Tasar", "left", "at", "2", "pm", "and", "reached", "Mughalpura", "after", "eight", "hours", "many", "people", "were", "killed", "on", "the", "way"] =22

Common unigrams between Reference Translation and Target Text1: Number of matching unigrams are 15 out of 22 in RT.

Unigram Precision: Precision = (Number of matching unigrams) / (Total unigrams in MTT1) P = 15/22 = 0.6818

BP = e^ (1 - 25/22) = 0.8725

BLEU-1 = P * BP = 0.5949 = 60%

c) Lingvanex

Lingvanex output is represented in tokenized form and evaluated to determine its closeness to the reference translation. Total tokens of lingvanex are 24.

["The", "special", "train", "from", "AmarTasar", "ran", "at", "two", "in", "the", "afternoon", "and", "reached", "Mughalpura", "after", "eight", "hours", "many", "men", "were", "killed", "on", "the", "way"] = 24

Number of matching unigrams is 15 out of 24

Precision = (Number of matching unigrams) / (Total unigrams in TT2)

P = 15/24 = 0.6250

BP = e ^ (1 - 25/24) = 0.9592

BLEU-1 = 0.5995= 60%

d) MateCat

The MateCat output is displayed in tokenized form and analyzed to examine its consistency with the reference translation. So, the total unigram tokens of Matecat are 20.

["A", "special", "train", "from", "AmarTasar", "reached", "Mughalpura", "at", "2", "pm", "after", "eight", "hours", "many", "people", "were", "killed", "on", "the", "way"] = 20

Number of matching unigrams: 10 out of 20

Precision = (Number of matching unigrams) / (Total unigrams in TT3)

P = 10/20 = 0.5

BP = e^ (1 - 25/20) = 0.778

BLEU-1 = 0.3894 = 39%

This suggests that the MT output has reasonable lexical accuracy, but there are some differences in word choice or content. TT3 introduces some major lexical and syntactic changes, which result in a lower match with the reference. These include different verb usage ("reached" vs. "left")

Machine Translation Tools Precision

TT1 (Google translate) 60%

TT2 (Lingvanex) 60%

TT3 (Matecat) 39%

Google Translate and lingvanex are faithful to the reference text in terms of word choice and meaning. They are both closer to reference translation in terms of BLEU metric analysis.

Qualitative Analysis

Various translation methods were used by the digital tools to render the text. The preservation of cultural terms like “Amritsar” and “Mughalpura” was an example of “borrowing”. The Urdu word " "دوپہرwas reduced in semantic scope when Google Translate and MateCat reduced its literal meaning to "afternoon." However, because the tools mainly concentrated on word-for-word rendering rather than contextual or cultural adaptation, literal translation was the predominant strategy across the output. These tools used the technique of literal translation for transferring source text to target text.

Analysis of Sentence # 2

This section analyses the second sentence of the story “Khol Do” by Sadat Hassan Manto.

SOURCE TEXT	REFERNCE TRANSLATION (HUMAN TRANSLATION)	GOOGLE TRANSLATE (TT1)	LINGVANEX (TT2)	MATECAT (TT3)
صبح دس بجے---کیمپ کی ٹھنڈی زمین پر جب سراج الدین نے آنکھیں کھولیں، اور اپنے چاروں طرف مردوں، عورتوں اور بچوں کا ایک متلاطم سمندر دیکھا تو اس کی سوچنے کی قوتیں اور بھی	At ten in the morning, when Sirajuddin opened his eyes on the ice-cold ground of the refugee camp, he saw a surging tide of men, women and children all around him and even his small remaining ability to think and comprehend	Ten o'clock in the morning. When Sirajuddin opened his eyes on the cold ground of the camp, and saw a turbulent sea of men, women and children around him, his powers of thought	Ten o'clock in the morning, When Sirajuddin opened his eyes on the cold ground of the camp, and saw a turbulent sea of men, women and children around him, his powers of thinking	At ten o'clock in the morning, When Sirajuddin opened his eyes to the cool land of the camp, and saw a desperate sea of men, women and children all around him, his thinking powers became even weaker

ضعیف ہو گئیں-	deserted him.	became even weaker.	became even weaker	

Quantitative Analysis The text has been analysed keeping in consideration Bleu metric analysis and its parameters.

a) Reference Translation

The following token count represents the reference translation, which serves as the baseline for comparison with the machine translation outputs. Total Tokens of reference translation are 43. ["At", "ten", "in", "the", "morning", "when", "Sirajuddin", "opened", "his", "eyes", "on", "the", "ice-cold", "ground", "of", "the", "refugee", "camp", "he", "saw", "a", "surging", "tide", "of", "men",", "women", "and", "children", "all", "around", "him", "and", "even", "his", "small", "remaining", "ability", "to", "think", "and", "comprehend", "deserted", "him".]

b) Google Translate

The Google Translate output is tokenized as follows for comparison with the reference translation. Total tokens of Google translate are 35.

["Ten", "o'clock", "in", "the", "morning", "When", "Sirajuddin", "opened", "his", "eyes", "on", "the", "cold", "ground", "of", "the", "camp",, "and", "saw", "a", "turbulent", "sea", "of", "men", "women", "and", "children", "around", "him", "his", "powers", "of", "thought", "became", "even", "weaker"]

TOTAL TOKENS: 35

$BP = e1 - 43/35 = e - 0.2286 \approx 0.7958$

$BLEU = 0.7958 \times 0.7714 = 0.6138$ or 61.38%

c) Lingvanex

The Lingvanex output is tokenized in the following manner for evaluation against the reference translation. Total tokens of Lingvanex are 36.

["Ten", "o'clock", "in", "the", "morning", "When", "Sirajuddin", "opened", "his", "eyes", "on", "the", "cold", "ground", "of", "the", "camp", "and", "saw", "a",

"turbulent", "sea", "of", "men", "women", "and", "children", "around", "him", "his", "powers", "of", "thinking", "became", "even", "weaker"]

TOTAL TOKENS: 36

BP = e1− 43/36 = e − 0.1944 ≈ 0.8235 0.8235 × 0.75 = 0.6176 or 61.76%

d) Matecat The MateCat output is broken down into tokens as shown below to facilitate BLEU score calculation. Total tokens of Matecat are 37.

["At", "ten", "o'clock", "in", "the", "morning", "When", "Sirajuddin", "opened", "his", "eyes", "to", "the", "cool", "land", "of", "the", "camp", "and", "saw", "a", "desperate", "sea", "of", "men", "women", "and", "children", "all", "around", "him", "his", "thinking", "powers", "became", "even", "weaker"]

BP = e1− 43/37 = e − 0.1622 ≈ 0.8504 0.8504 × 0.7027 = 0.5976 or 59.76%.

MT Tools Precision

Google Translate 61%

Lingvanex 62%

Matecat 60%

Here in this sentence, Lingvanex remains closer to the reference translation by securing 62% accuracy in BLEU analysis.

Qualitative Analysis

Google Translate, Lingvanex and MateCat all employ literal translation and closely follow the original text for sentence structure. For instance, all three preserve “Ten o'clock in the morning” and “saw a turbulent sea of men, women and children”. Google Translate, for example, uses reduction and simplifies the metaphorically rich phrase “even his small remaining ability to think and comprehend deserted him” to “his powers of thought became even weaker”. Similarly, Lingvanex also applies reduction, while additionally using modulation, replacing “powers of thought” with “powers of thinking”, slightly shifting the grammatical structure. While MateCat demonstrates these substitution and lexical

modulation when referring to "desperate sea" instead of "surging tide" and "cool land" for "ice-cold ground" losing some emotional touch. Overall, despite all three tools straying towards literal translation, they also utilize specific translation techniques like reduction, modulation, and substitution which influence precision and expressiveness in a diverse manner.

Analysis of Sentence # 3.

This section analyses the third sentence of the story "Khol Do" by Sadat Hassan Manto.

SOURCE TEXT	REFERENCE TRANSLATION (HUMAN TRANSLATION)	GOOGLE TRANSLATE (TT1)	LINGVANEX (TT2)	MATECAT (TT3)
وہ دیر تک گدلے آسمان کو ٹکٹکی باندھے دیکھتا رہا، یو تو طرف شور برپا تھا، لیکن بوڑھے سراج الدین کے کان جیسے بند تھے ، اسے کچھ سنائی نہیں دیتا تھا، کوئی اسے دیکھتا تو یہ خیال کرتا تھا، کہ وہ	He stared at the murky sky for the longest time. Amidst the incredible din, his ears seemed to be firmly plugged against any sound. Seeing him in this state anyone would have thought he was deeply engrossed in thought. That, of course, was not the case. He was totally numb. His entire	He stared at the cloudy sky for a long time. There was noise everywhere in the camp, but old Sirajuddin's ears were closed; he could not hear anything. When one saw him, one would think	For a long time, he watched the muddy sky tied with a ticket; there was noise everywhere in the camp, but old Sirajuddin's ears were closed, he could not hear anything. When someone saw	He watched the muddy sky for a long time; there was noise everywhere in the camp, but old Sirajuddin's ears were closed, he could not hear anything.

کسی گہری فکر میں غرق ہے ، اسے کچھ سنائی نہیں دیتا ہے ، اس کے ہوش و حواس شل تھے ، اس کا سارا وجود خلا میں معلق تھا	being seemed to be suspended in space.	that he was immersed in some deep thought; he could not hear anything, his senses were paralyzed, his whole being was suspended in space.	him, they thought he was immersed in a deep worry; he did not hear anything, his senses were paralyzed, his whole being was suspended in space.	When someone saw him, they thought he was immersed in some deep thought; he did not hear anything, his senses were numb, his whole existence was suspended in space.

Quantitative Analysis The text has been analysed keeping in consideration Bleu metric analysis and its parameters.

a) Reference Translation

The following token count represents the reference translation, which serves as the baseline for comparison with the machine translation outputs. Total Tokens of reference translation are 59. ["He", "stared", "at", "the", "murky", "sky", "for", "the", "longest", "time", "Amidst", "the", "incredible", "din", "his", "ears", "seemed", "to", "be", "firmly", "plugged", "against", "any", "sound", "."Seeing", "him", "in", "this", "state", "anyone", "would", "have", "thought", "he", "was", "deeply", "engrossed", "in", "thought, "That", "of", "course", "was", "not", "the",

"case", "He", "was", "totally", "numb", "His", "entire", "being", "seemed", "to", "be", "suspended", "in", "space".] Total Tokens: 59

b) Google Translate:

The Google Translate output is tokenized as follows for comparison with the reference translation. Total Tokens of Google Translate are 50.

["He", "stared", "at", "the", "cloudy", "sky", "for", "a", "long", "time", "There", "was", "noise", "everywhere", "in", "the", "camp", "but", "old", "Sirajuddins", "ears", "were", "closed", "he", "could", "not", "hear", "anything", "When", "one", "saw", "him", "one", "would", "think", "that", "he", "was", "immersed", "in", "some", "deep", "thought", "he", "could", "not", "hear", "anything", "his", "senses", "were", "paralyzed", "his", "whole", "being", "was", "suspended", "in", "space"]

Total tokens: 59

Candidate length (c) = 59

Matching unigrams = 19

Precision = 59/19 ≈ 0.3220

BP = 1

BLEU = 1 × 0.3220 = 0.322

c) Lingvanex

The Lingvanex output is broken down into tokens as shown below to facilitate BLEU score calculation. Total Tokens of Lingvanex translation are 62.

["For", "a", "long", "time", "he", "watched", "the", "muddy", "sky", "tied", "with", "a", "ticket", "there", "was", "noise", "everywhere", "in", "the", "camp", "but", "the", "old", "Sirajuddins", "ears", "were", "closed", "he", "could", "not", "hear", "anything", "when", "someone", "saw", "him", "he", "thought", "that", "He", "is", "immersed", "in", "a", "deep", "worry", "he", "does", "not", "hear", "anything",

"his", "senses", "were", "paralyzed", "His", "whole", "being", "was", "suspended", "in", "space"]

Total Tokens: 62

Candidate length (c) = 62

Matching unigrams = 12

P = 62/12 ≈ 0.1935 BP = 1

BLEU = 1×0.1935 = 0.194

d) Matecat:

The Matecat translation tokens are 57 as shown below.

["He", "watched", "the", "muddy", "sky", "for", "a", "long", "time", "there", "was", "noise", "everywhere", "in", "the", "camp", "but", "the", "old", "Sirajuddins", "ears", "were", "closed", "he", "could", "not", "hear", "anything", "when", "someone", "saw", "him", "he", "thought", "that", "he", "was", "immersed", "in", "some", "deep", "thought", "he", "did", "not", "hear", "anything", "he", "had", "senses", "his", "whole", "existence", "was", "suspended", "in", "space"] Total tokens: 57

Candidate length (c) = 57

Matching unigrams = 16

P = 57/16 ≈ 0.2807

BP= e (1−59/57) = e − 0.0351 ≈ 0.9655

BLEU = 0.9655 × 0.2807 ≈ 0.271

MT Tools Precision

Google Translate 32%

Lingvanex 19%

Matecat 28%

Qualitative Analysis

In this segment, Google Translate uses a mix of literal translation and modulation, evident in phrases like "cloudy sky" replacing "murky sky", and "immersed in deep thought" echoing the source meaning but with slightly altered imagery. Google Translate got the highest BLEU score among the three tools and its higher unigram match suggests a relatively faithful rendering. Lingvanex, however, demonstrates semantic mistranslation, with its literal translation gone wrong, such as "tied with a ticket" for " ,"ٹکٹکی باندھےwhich reflects a failure in idiomatic equivalence and shows poor application of adaptation or borrowing. MateCat, while somewhat more accurate than Lingvanex, applies modulation and literal translation, as

70 seen in "existence" for "being" and "immersed in some deep thought", which shows a stronger attempt at preserving meaning, though it slightly deviates lexically. Overall, Google Translate aligns most closely with the reference in both lexical choices and translation strategies.

Analysis of Sentence# 4

This section analyses fourth sentence of the story “Khol Do” by Sadat Hassan Manto.

SOURCE TEXT (URDU)	**REFERENCE TRANSLATION (HUMAN TRANSLATION)**	**GOOGLE TRANSLATE (TT1)**	**LINGVANEX (TT2)**	**MATECAT (TT3)**
گدلے آسمان کی طرف	Gazing blankly at the murky sky, his	Looking aimlessly at the murky	Seeing the muddy sky without	Looking at the sky

بغیر کسی ارادے کے دیکھتے دیکھتے سراج الدین کی نگاہیں سورج سے ٹکرائیں، تیز روشنی اس کے وجود کے رگ و ریشے میں اتر گئی اور وہ جاگ اٹھا، اوپر تلے اس کے دماغ پر کئی تصویریں دوڑ گئیں، لوٹ، آگ۔۔۔۔بھاگم بھاگ۔۔۔۔اسٹیشن۔۔۔گو لیاں۔ ۔۔رات اور سکینہ سراج الدین ایک دم اٹھ کھڑا ہوا اور پاگلوں کی طرح اس نے اپنے چاروں طرف	eyes collided with the sun and a shaft of intense light penetrated every fiber of his being. Suddenly he snapped back into consciousness. A series of images flitted across his mind plunder, fire, stampede, the train station, gunshots, night, and Sakina. Sirajuddin jumped up with a start and made	sky, Sirajuddin's eyes met the sun. The bright light fell into the veins and fibers of his being, and he woke up. Several images flashed through his mind loot, fire, runaway, station, bullets, night, and sky. Sirajuddin stood up at once, and like a madman, he began	any intention, Sirajuddin's eyes hit the sun. The bright light fell into the veins and fibers of his existence, and he woke up. Many pictures ran through his mind loot, fire, run away, station, tablets, night, and Sakina. Sirajuddin immediately stood up and, like	without any intention, Sirajuddin 's eyes hit the sun. The bright light descended into the veins of his being, and he woke up. Many images ran through his mind from above loot, fire, runaway, station, bullets, night, and

پھلیے ہوئے انسانوں کے سمندر کو کھٹکانا شروع کیا۔	his way through the interminable tide of humanity around him like a man possessed.	to knock over the sea of men that had swarmed around him.	a madman, started knocking the sea of humans spread around him.	Sakina. Sirajuddin got up at once and, like a madman, began to shake the sea of men around him.

Quantitative Analysis The text has been analysed keeping in consideration Bleu metric analysis and its parameters.

a) Reference Translation

The tokenization of the reference translation serves as the standard text against which the machine translation outputs are evaluated. Reference Translation tokens are 72 in this sentence.

["Gazing", "blankly", "at", "the", "murky", "sky", "his", "eyes", "collided", "with", "the", "sun", "and", "a", "shaft", "of", "intense", "light", "penetrated", "every", "fiber", "of", "his", "being", "Suddenly", "he", "snapped", "back", "into", "consciousness", "A", "series", "of", "images", "flitted", "across", "his", "mind", "images", "of", "plunder", "fire", "stampede", "the", "train", "station", "gunshots", "night", ",", "and", "Sakina", "Sirajuddin", "jumped", "up", "with", "a", "start", "and", "made", "his", "way", "through", "the", "interminable", "tide", "of", "humanity", "around", "him", "like", "a", "man", "possessed".]

Total Tokens: 72

b) Google Translate

The tokenization of the Google Translate output is presented below for comparison with the reference translation. Here, in this sentence, total tokens of GT are 62.

["Looking", "aimlessly", "at", "the", "murky", "sky", "Sirajuddin's", "eyes", "met", "the", "sun", "the", "bright", "light", "fell", "into", "the", "veins", "and", "fibers", "of", "his", "being", "he", "woke", "up", "several", "images", "flashed", "through", "his", "mind", "loot", "fire", "runaway", "station", "bullets", "night", "and", "sky", "Sirajuddin", "stood", "up", "at", "once", "like", "a", "madman", "he", "began", "to", "knock", "over", "the", "sea", "of", "men", "that", "had", "swarmed", "around", "him]

Total Tokens (GT): 62

Matching = 25 c = 68, r = 72

Precision = 25 / 62 ≈ 0.403

BP = exp (1 - 72 / 62) exp (− 0.16129) ≈ 0.851

BLEU = Precision × BP = 0.403 × 0.851 ≈ 0.343

c) Lingvanex

The tokenization of the Lingvanex output presented the total tokens as 59.

["Seeing", "the", "muddy", "sky", "without", "any", "intention", "Sirajuddin's", "eyes", "hit", "the", "sun", "the", "bright", "light", "fell", "into", "the", "veins", "and", "fibers", "of", "his", "existence", "he", "woke", "up", "many", "pictures", "ran", "down", "his", "mind", "loot", "fire", "run", "away", "station", "tablets", "night", "and", "Sakina", "Sirajuddin", "immediately", "stood", "up", "like", "a", "madman", "he", "started", "knocking", "the", "sea", "of", "humans", "spread", "around", "him"]

Total Tokens (Lingvanex): 59

Matching = 22 c = 59

Precision = 22 / 59 ≈ 0.373 BP = exp (1 - 72 / 59) ≈ exp (-0.220) ≈ 0.802

BLEU = 0.373 × 0.802 ≈ 0.299

d) MateCat

The MateCat output is broken down into tokens as shown below to evaluate BLEU score calculation, while the total tokens are 62.

["Looking", "at", "the", "mattress", "without", "any", "intention", "towards", "the", "sky", "Sirajuddin's", "eyes", "hit", "the", "sun", "the", "bright", "light", "descended", "into", "the", "vein", "of", "his", "existence", "he", "woke", "up", "many", "pictures", "ran", "on", "his", "mind", "from", "above", "loot", "fire", "runaway", "station", "tablets", "night", "and", "Sakina", "Sirajuddin", "got", "up", "at", "once", "like", "a", "madman", "he", "began", "to", "shake", "the", "sea", "of", "men", "around", "him"]

Total Tokens (MateCat): 62

Matching = 21 c = 62

Precision = 21 / 62 ≈ 0.339

BP = exp (1 - 72 / 62) ≈ exp (-0.161) ≈ 0.851

BLEU = 0.339 × 0.851 ≈ 0.288

Mt Tools Precision

Google Translate 34%

Lingvanex 30%

Matecat 29%

Qualitative Analysis

In this instance, Google Translate achieves the highest BLEU score due to greater lexical alignment with the reference, despite syntactic differences. It applies literal translation for phrases like “bright light”, “his mind”, and “around him”, and shows use of modulation in "met the sun" vs. “collided with the sun.” Google Translation also uses the technique of borrowing where it borrowed the word

"loot", whose literal meaning is plunder. Lingvanex introduces slight deviations, such as "hit the sun" and "pictures ran down his mind," reflecting transposition but still retains a solid lexical overlap. MateCat, while semantically close, suffers from critical lexical errors such as "shake" instead of "make his way," indicating lexical creation and inaccurate equivalence. It also inserts "from above", which doesn't exist in the original, reflecting amplification. Thus, while all three convey the narrative flow, Google Translate performs best in BLEU terms here.

Analysis of Sentence # 5

This section analyses fifth sentence of the story "Khol Do" by Sadat Hassan Manto.

SOURCE TEXT (URDU)	**REFERENCE TRANSLATION (HUMAN TRANSLATION)**	**GOOGLE TRANSLATE (TT1)**	**LINGVANEX (TT2)**	**MATECAT (TT3)**
پورے تین گھنٹے وہ سکینہ سکینہ پکارتا کیمپ کی خاک چھانتا رہا، مگر اسے اپنی جوان اکلوتی بیٹی کا پتہ نہ چل سکا،چاروں طرف ایک دھاندلی سی مچی ہوئی تھی، کوئی اپنا بچہ ڈھونڈ رہا تھا، کوئی ماں، کوئی بیوی اور کوئی بیٹی، سراج الدین تھک	For three full hours he scoured the camp calling out "Sakina! Sakina!" but found no trace of his only daughter, a teenager. The whole area was rife with ear- splitting noise. Someone was looking for his child, another for his mother, still another for his wife or	For three whole hours he called "Sakina! Sakina!" and searched the dust of the camp, but he could not find his young and only daughter. There was chaos all around. Someone was looking for his child, some for mother, some for wife and some	For three whole hours, he kept sifting through the dust of the camp while calling "Sakina! Sakina!", but he could not find his young only daughter. There was chaos all around— someone was looking for his child, a mother, a wife,	For the whole three hours, he continued to search the camp calling "Sakina! Sakina!" but could not find his young only daughter. There was chaos around— someone looking for his child, a mother, a

ہار کر ایک طرف بیٹھ گیا، اور حافظے پر زور دینے لگا، کہ سکینہ اس سے کب اور کہاں جدا ہوئی تھی لیکن سوچتے سوچتے اس کا دماغ سکینہ کی ماں کی لاش پر جا کر جم جاتا ہے ، جس کی ساری انتڑیاں باہر نکلی ہوئی تھیں، اس سے آگے وہ اور کچھ نہ سوچ سکا	daughter. Finally, Sirajuddin gave up and plopped down off to one side from sheer exhaustion, straining his memory to recall when and where Sakina had become separated from him. However, each effort to recall ended with his mind jammed at the sight of his wife's mutilated body, her entrails spilling out, and he couldn't think any further.	for daughter. Sirajuddin sat down tiredly and started pressing his hafiza when and where Sakina was separated from him. But thinking of it, his mind went to the body of Sakina's mother, whose entrails were out; he could not think of anything else.	a daughter. Sirajuddin grew tired and sat to one side, trying to recall when and where Sakina was separated from him. But while thinking, his mind went to the dead body of Sakina's mother and froze; her entrails were protruding out, and he could think of nothing else.	wife, a daughter Sirajuddin sat down exhausted, insisted to hafeez when and where Sakina had separated from him. But while thinking, his mind drifted to Sakina's mother's body, all her intestines spilling out; he could not think of anything more.

Quantitative Analysis

The text has been analysed keeping in consideration Bleu metric analysis and its parameters.

a) Reference Translation

The tokenization of the reference translation is provided first, as it serves as the standard text against which the machine translation outputs are evaluated. Total tokens of reference translation are 103.

[For", "three", "full", "hours", "he", "scoured", "the", "camp", "calling", "out", "Sakina", "!", "Sakina","but", "found", "no", "trace", "of", "his", "only",

"daughter", "a", "teenager", "The", "whole", "area", "was", "rife", "with", "ear-splitting", "noise", "Someone", "was", "looking", "for", "his", "child", "another", "for", "his", "mother", "still", "another", "for", "his", "wife", "or", "daughter", "Finally", "Sirajuddin", "gave", "up", "and", "plopped", "down", "off", "to", "one", "side", "from", "sheer", "exhaustion", "straining", "his", "memory", "to", "retrieve", "the", "moment", "when", "Sakina", "had", "become", "separated", "from", "him". "However", "each", "effort", "to", "recall", "ended", "with", "his", "mind", "jammed", "at", "the", "sight", "of", "his", "wife's", "mutilated", "body", "her", "guts", "spilling", "out", "and", "he", "couldn't", "go", "any", "further''.]

Total Tokens: 103

b) Google Translate:

The tokenization of the Google Translate output is presented below for comparison with the reference translation, while the tokens are 86. ["For","three","whole","hours","he","called","sakina","sakina","and","searched", "the","dust",

"of","the","camp","but","he","could","not","find","his","young","and","only","daughter",

"There","was","a","rigging",all",around","someone","was","looking","for","his","child","some","mother","some","wife","and","some","daughter","sirajuddin","sat","down","tiredly","and","stated","stressing","on","hafiza","when","and","where","sakina","was","separated","from","him",

"but","thinking","about","it","his","mind","went","to","the,"body","of","sakina's","mother"," whose","entrails","were","all","out","he","could","not","think", "of", "anything", "else".]

Total Tokens: 86

Matching Unigrams: 47

P= 47 / 86 = 0.5465

BP = e (1−103/86) = exp (1−1.1977) = exp (−0.1977) ≈ 0.8207

BLEU: 0.8207 × 0.5465 ≈ 0.4486

b) Lingvanex

Lingvanex target text is tokenized in the following manner for scoring against the reference translation. Here, total tokens are 95.

['For", "three", "whole", "hours", "he", "kept", "sifting", "the", "dust", "of", "the", "camp", "while", "calling", "Sakina", "Sakina", "but", "he", "could", "not", "find", "his", "young", "only", "daughter", "there", "was", "a", "rigging", "all", "around", "someone", "was", "looking", "for", "his", "child", "a", "mother", "a", "wife", "and", "a", "daughter", "Sirajuddin", "got", "tired", "and", "sat", "on", "one", "side", "and", "started", "emphasizing", "memory", "when", "and", "where", "Sakina", "was", "separated", "from", "him", "but", "thinking", "about", "it", "her", "mind", "goes", "to", "the", "dead", "body", "of", "Sakina's", "mother", "and", "freezes", "all", "the", "entrails", "of", "which", "were", "protruding", "out", "she", "could", "not", "think", "of", "anything", "else".]

Total Tokens: 95

P= 51/95= 0.5368

BP = e (1−103/95) = e − 0.0842 ≈ 0.9193

BLEU = 0.9193 × 0.5368 ≈ 0.4935

c) Matecat

After transferring source text into target text from Matecat, the total words are given individual tokens. So, the total tokens of Matecat are 81.

["For", "the", "whole", "three", "hours", "he", "continued", "to", "scrutinize", "the", "camp", "calling", "Sakina", "Sakina", "but", "he", "could", "not", "find", "his", "young", "only", "daughter", "there", "was", "a", "rigging", "around", "someone", "was", "looking", "for", "his", "child", "a", "mother", "a", "wife", "and", "a", "daughter", "Sirajuddin", "tiredly", "sat", "on", "one", "side", "and",

"insisted", "on", "Hafeez", "when", "and", "where", "Sakina", "had", "separated", "from", "him", "but", "thinking", "his", "mind", "goes", "to", "Sakina's", "mother's", "body", "all", "of", "whose", "intestines", "were", "out", "he", "could", "not", "think", "of", "anything", "else"]

Total Tokens: 81

$P = 47 / 81 = 0.5802$

$BP = e\,(1-103/81) = e - 0.2716 \approx 0.7622$

$BLEU = 0.7622 \times 0.5802 \approx 0.4423$

Mt Tools Precision

Google Translate 45%

Lingvanex 50%

Matecat 44%

Qualitative Analysis

Within this segment, Lingvanex not only achieves the highest number of matching unigrams with 51 but also preserves the context more faithfully than others. It uses modulation by maintaining fluency like in “he kept sifting the dust of the camp” instead of “he scoured the camp.” Moreover, transposition is also used in rearranging the grammatical categories. With that said, there is a minimal semantic discrepancy results from linguistic amplification of “her mind for Sirajuddin”, which creates a gender inconsistency. Google Translate having 47 matching unigrams suffered from the issue of being overly literal, yielding “stress on Hafiza,” which is a phrase that does not make sense. Such an utterance illustrates a failed literal transfer as well as interference from source syntax. MateCat also with 47 matches shows established equivalent in “scrutinize” for “scoured,” but it suffers from distortion with erroneous substitutions like “insisted on Hafeez.” Overall, Lingvanex manages the best balance between semantic retention and natural expression in target language by means of modulation and transposition,

while GT and MateCat tend towards direct transfer, which results in a lack of clarity of meaning.

5. Results and Discussion

Quantitative analysis showed that the Google Translate had the highest BLEU unigram precision scores compared to the other three machine translation (MT) tools, which indicated the highest lexical similarity to the reference translation by M. Umar Memon. Lingvanex was second, and MateCat produced the lowest BLEU scores due to lower unigram precision and higher token mismatches. BLEU proved to be useful in terms of the lexical correspondence between the machine-generated and reference translation, but it did not measure the stylistic, semantic or cultural aspects of literary translation.

The qualitative analysis based on the translation techniques proposed by Molina and Albir (2002) showed that the three MT systems used literal translation as the dominant technique, which indicated a high preference for formal equivalence. In general, Google Translate preserved semantic meaning better, and sometimes used borrowing to preserve culturally specific terms. Lingvanex did not depend much on amplification and often gave awkward or incorrect translations of idiomatic phrases. The worst results were achieved by MateCat, which generated many literal translations and calques, frequently distorting the meaning and reducing textual coherence. The results demonstrated that while current MT systems were able to produce lexical content with a fair degree of accuracy, they were limited in their ability to preserve stylistic richness, emotional depth and cultural nuances of literary texts.

The quantitative and qualitative results combined indicated that BLEU scores are not sufficient for evaluating the quality of literary translation. Automatic metrics gave an objective measure of lexical accuracy, but they did not measure literary style, contextual meaning and cultural fidelity. Therefore, the combination

of BLEU evaluation and translation technique analysis provided a more holistic evaluation of machine translation performance. These findings were in line with earlier research indicating that Google achieved better lexical accuracy, but also reaffirmed the continued dependence of machine translation systems on literal translation strategies in the literary domain.

In general, the above evaluation showed that Google Translate performed better than Lingvanex and MateCat in translating Khol Do from Urdu to English. However, all of the aforementioned machine translation tools were not able to completely convey the cultural and stylistic richness of the literary language used by Manto. According to the study results, machine translation can be considered as a very good tool in generating preliminary translations. However, it cannot substitute human translators in the translation of culturally and stylistically complex literary work. This study was one of the first to incorporate BLEU unigram precision together with the criteria established by Molina and Albir into Urdu–English literary machine translation evaluation framework**.**

6. Conclusion

This study analyzed the efficiency of Google Translate, Lingvanex, and MateCat when translating Saadat Hasan Manto's Khol Do from Urdu to English by using the BLEU score based on unigram precision as well as Molina and Albir's classification of translation techniques. Quantitatively, it was found that Google Translate had significantly higher BLEU scores, implying closer lexical match to the reference translation. However, Lingvanex and MateCat had lower scores because of lexical and structural mismatches. Qualitatively, all three tools used mostly literal translation and applied techniques such as borrowing and reduction to a very limited extent. Even though Google Translate maintained the integrity of the source text better than the other tools, none of them managed to reflect the style, culture, idiom, and emotions of the original work.

The results indicated that while existing MT systems were helpful in understanding Urdu literary texts, they did not suffice for creating high quality literary translations without human assistance. The combination of automated BLEU scoring with translation approaches to evaluation performed in this research allowed for a comprehensive assessment of MT systems efficiency in translating Urdu literary texts into English language. Future researchers can use this study as a foundation for exploring other machine translation tools, texts, and evaluation methods.

References

Ayesha, M. A., Noor, S., Ramzan, M., & Khan, H. U. (2017). Evaluating Urdu to Arabic machine translation tools. *International Journal of Advanced Computer Science and Applications, 8*(10).

Al-Khresheh, M. H., & Almaaytah, S. A. (2018). English proverbs into Arabic through machine translation. *International Journal of Applied Linguistics and English Literature, 7*(5), 158–166. https://doi.org/10.7575/aiac.ijalel.v.7n.5p.158

Ali, M. A. (2020). Quality and machine translation: An evaluation of online machine translation of English into Arabic texts. *Open Journal of Modern Linguistics, 10*(5), 524–548. https://doi.org/10.4236/ojml.2020.105030

Catford, J. C. (1965). A linguistic theory of translation.

Chakrawarti, R. K., Mishra, H., & Bansal, P. (2017). Review of machine translation techniques for idea of Hindi to English idiom translation. *International Journal of Computational Intelligence Research, 13*(5), 1059–1071. Doherty, S. (2016). Translations| the impact of translation technologies on the process and product of translation. *International journal of communication*, *10*, 23.

Hatim, B., & Munday, J. (2019). *Translation: An advanced resource book for students*. Routledge. Hutchins, J. (2005). The history of machine translation in a nutshell. *Retrieved December*, *20*(2009), 1-1.

Sipayung, K. T., Sianturi, N. M., Arta, I. M. D., Rohayati, Y., & Indah, D. (2021). Comparison of translation techniques by Google Translate and U-Dictionary: How differently do both machine translation tools perform in translating? *Elsya: Journal of English Language Studies, 3*(3), 236–245. https://doi.org/10.31849/elsya.v3i3.6954

Stevanović, I., & Radičević, L. (2012). Comparative analysis of machine translation systems. *International Journal of Computer Applications, 12*(2), 5–8. https://doi.org/10.5120/ijca2012553

Yuliasri, I. (2016). Translation techniques and pragmatic equivalence in Indonesian translation of humorous utterences in the Walt Disney's Donald duck comics. In *Prasasti: Conference Series* (pp. 409-414). Zong, Z. (2018, September). Research on the relations between machine translation and human translation. In *Journal of Physics: Conference Series* (Vol. 1087, No. 6, p. 062046). IOP Publishing.