Available online at www.garmian.edu.krd

Passer Journal

Passer 6 (Issue 2) (2024) 351-368

http://passer.garmian.edu.krd/

A Linguistic Evaluation of Google Translate Between Human and Machine Translation for English-Sorani Kurdish Languages Using Natural Language Processing (NLP)

Dlnya Raouf Muhealddin[1*], Yusra Mohammed Mohammed Salih[2], Fatima Jalal Taher[3]

[1] *International English Language Center, Sulaimani Polytechnic University, Sulaymaniyah, Sulaymaniyah 46001, Kurdistan Region, Iraq.*
[2]*Department of Computer Engineering, Faculty of Engineering, Tishk International University, Sulaymaniyah 46001, Kurdistan Region, Iraq.*
[3]*Department of Translation, College of Languages, University of Sulaimani, Sulaymaniyah 46001, Iraq.*

Received 27 October 2023; revised 22 May 2024;
accepted 22 May 2024; available online 23 June 2024

DOI: 10.24271/PSR.2024.422619.1415

ABSTRACT

Online machine translation systems are widespread worldwide and free of charge. Among the most widely utilized multilingual machine translation services is Google Translate; however, Sorani Kurdish was one of the languages added to Google Translate in May 2022. The assessment of machine translation systems is crucial for understanding their performance. Several standard and efficient natural language processing metrics accurately approximate human assessments, including BLEU, NIST, and BERTScore. This study is interdisciplinary research between two intertwined fields: language and technology. In computational linguistics, Sorani Kurdish has received trivial attention; therefore, the current work targets to investigate and evaluate sets of English to Sorani Kurdish translations to determine whether machine translations differ from human translations. The machine translation was analyzed from a descriptive quantitative and qualitative perspective according to the investigators' mother language intuitions (Sorani Kurdish) and their linguistic backgrounds of English and Kurdish regarding morphological, semantic, and syntactic features. Based on the results of these experiments, Google Translate performs at a lower level than the average performance of Kurdish native speakers, and further research is suggested to increase its quality. Despite several shortcomings in evaluating Google Translate utilizing NLP and the ongoing advances of Google Translate, Sorani Kurdish speakers can gain a lot from using this free online machine translation service.

Keywords: Google Translate (GT), English-Kurdish Translation, Kurdish-Sorani Dialect, Linguistic Evaluation, Natural Language Processing (NLP), Machine Translation (MT), BLEU Score, NIST, BERTScore.

1. Introduction

Translation is one of the areas that has profoundly benefited from the expansion of technology and AI because it is a vital and highly needed profession. The past advancements in technology and the progress of AI have affected many aspects and made significant changes in all the standards. Translation was one of the majors that began with only human assistance. However, as new technological tools emerged, they occasionally switched to machine-assisted human translation. Recently, machine translation has had a significant impact on translation. As a result, the once-barely used term "translation" has transformed into a "machine translation system," which can be utilized for various purposes with some limitations. In addition, machine translation systems compromise between effectiveness and efficiency. However, machine translations are less accurate and intelligible than human translations, but they are quicker and simpler [1].

* Corresponding author
E-mail address: dlnya.raouf@spu.edu.iq (Instructor).
Peer-reviewed under the responsibility of the University of Garmian.

Kurdish is an Indo-European language with four dialects, as shown in Figure 1. The Kurdish-Sorani dialect uses Arabic script, while Kurmanji uses Latin script. Hossani and Medijodiyec [2] declared that Kurds in areas of Iraq and Iran speak the Central Kurdish dialect known as Sorani. Despite having a different sound from Arabic, Sorani uses the Arabic alphabet to write many of its words. Meanwhile, English is a Germanic language and belongs to the Western Germanic branch; it uses Latin script for writing. As Miller stated, English is alphabetic, like many other European languages, such as Spanish, German, French, and Italian. Despite differences in many aspects, syntactically, morphologically, and semantically, the orthography of both languages is of two very different types [3]. So, the evaluation will be done based on linguistic criteria and the metrics of NLP used to aid in accurate evaluation.

Additionally, according to [5], the Kurdish language is categorized within the Indo-Iranian branch of the Indo-European language family.

While research in computational linguistics can be qualitative,

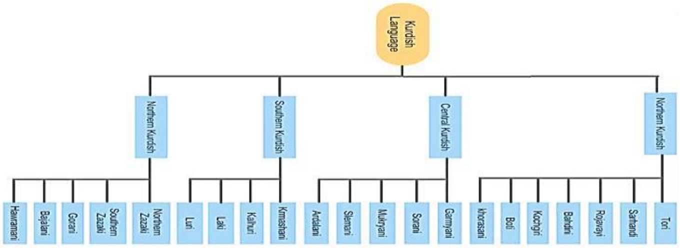


Figure 1: Kurdish dialects and sub-dialects diagram [4].

with the dataset and corpus of machine translation, it can also be quantitative. In this way, machine translation (henceforth MT) is carried out automatically by locating translation units from the source language in the target language. MT may discover the correct connotation or numerous meanings of a translation unit. However, it may need to include the appropriate usage or sense of idiomatic expressions and polysemous terms in a given context. Because of this, translations are frequently mechanical rather than creative or communicative, i.e., the fundamental building block of the "art" of translation is often absent. Professional translators apply their trade to produce the final translation by finding the appropriate meaning using their abilities, knowledge, and experience.

Automatic machine translation evaluation (MTE) is one of the most essential applications of natural language processing (NLP). It has been studied to substitute human evaluation in machine translation development because it is low-cost, handy, and stable [6]. Multiple commonly utilized and effective natural language processing metrics, including BLEU, NIST, and BERTScore, provide accurate approximations of human assessments. All these metrics need reference translations because they compare the MT output with reference translations (human translation, henceforth HT) and provide comparison scores.

Machine translation services such as Google Translate (GT) are widely used globally, and many languages have been translated using GT. Google Translate allows users to translate texts, documents, and websites between multiple languages using neural machine translation, a technology developed by Google [6]. Additionally, Google Translate claims to serve around 500 million whole consumers daily, with over 100 billion translated words every day [7]; as of September 2023, 133 languages are supported at various levels, and the number of languages supported by Google Translate has increased gradually. Kurmanji Kurdish first appeared on Google Translate in 2016, whereas Sorani Kurdish was added in 2022. Because many Sorani Kurdish speakers frequently use Google Translate, developing Sorani Kurdish on Google was crucial. However, a machine cannot fully replace a human, which is valid for translation. Human translators can find translated words, texts, and sentences in the target languages quickly and accurately using their translation skills. In contrast, a machine can use algorithms to locate the source language's equivalent in the target language, and algorithms are created to analyze sentences [8].

In this study, we applied the evaluation metrics BLEU, NIST, and BERTScore to assess and compare a machine translation system that relies on statistical approaches for translating Sorani Kurdish and English. Specifically, we analyzed Google Translate's performance, comparing it with human translation based on N-grams. Since the inclusion of Sorani Kurdish in Google Translate proved highly beneficial for the concept we wish to present in our paper, this endeavor strives to be at the forefront of research in this area, mainly focusing on Machine Translation Systems (MTS). We also introduce a novel human evaluation method conducted by linguistic experts, which involves categorizing translation errors and strengths into four levels: morphological, lexical, semantic, and syntactic.

The number of errors identified at each level is then used to compare human and machine translations. The primary goal is to evaluate the linguistic accuracy of Google Translate, which is linguistically supported by natural language processing. Furthermore, we aim to determine which linguistic errors are more likely to be discovered by human analyzers.

To test the metrics, we arranged a corpus containing words, compound words, sentences, and paragraphs with their Kurdish machine translations and their translation obtained from human translation (native speakers' intuition, the first and third authors). The dataset is divided into three separate corpora as the source dataset: the English language dataset, the machine translation-MT (candidate) dataset, and the human translation-HT (reference) dataset. The machine translation dataset contains the English language dataset's correspondent translation, accessed from the openly accessible Google Translate (Google, https://translate.google.com) English-Kurdish (Sorani) translation machine tool. The HT (reference) dataset also contains the correspondent translation for the English language dataset. The dataset is tested using Python to evaluate the metrics. The dataset is imported into the program. Then, random data is selected from the dataset and given the algorithm to calculate the scores and compare the results with HT (reference).

The findings imply that, for the most part, Google Translate is reliable and substantially accurate. At the same time, some inaccuracies may be picked up in the study utilizing evaluation metrics or in certain rare instances where they may not be. This platform may be trusted based on morphology because it is generally correct for single and compound words. Syntactically, it can be trusted for sentences and paragraphs, but morphosyntactically, it may need to be more accurate. In terms of semantics, it can be trusted with straightforward content and sentences. However, cultural differences between the two languages make them less reliable with complex ones, such as idioms, proverbs, and common expressions. Despite several shortcomings found in evaluating Google Translate utilizing natural language processing and the ongoing advances of Google Translate, Sorani speakers of Kurdish can gain a lot from this machine translation service.

Finally, given the rise of Sorani Kurdish and the popularity of Google Translate among Sorani Kurdish speakers, it is crucial to assess the quality of Google Translate and help the regulators of the Sorani dialect of the Kurdish language on Google Translate to review their work where needed and strengthen their abilities through linguistic analysis in both Kurdish and English. This effort also shows and appreciates their work and indicates the significance of having Sorani Kurdish as one of the languages found on Google Translate so that speakers can access accurate information and resources online.

2. Literature Review

In the past, translation acts were performed by humans. However, with the advance of technology, translation has become one of the areas that has seen significant advantages from technology. Therefore, machine translation systems have become one of this era's most frequently used applications and tools. Nowadays, many people of different specializations depend on MTS for translation. Examples of machine translation systems are Bing, Babylon, Collins, inKurdish, and Google Translate, which are used to translate various languages worldwide among the current systems. Past research findings have shown a preference for Google Translate over other machine translation tools. In 2006, Google Translate was launched and is now considered one of the most popular machine translation services [8]. Although Google Translate offers translations between various languages, the accuracy varies considerably.

Thus, according to [1], for more than forty years, computer scientists and linguists have dedicated their research to diagnosing the weaknesses of machine translators and creating systems that will perform more quickly and with higher accuracy. Today, machine translation has been regarded as an acceptable means to garner the "gist" of a given text using evaluation metrics (such as BLEU, NIST, and METEOR) that could be taken to evaluate the translated corpus [9-10], even though it cannot yet generate a translation whose accuracy and quality are equal to that of a human translator.

Aiken [7] performed a study and assessed the translation quality of Google Translate across 50 languages rather than only two languages. According to his findings, translations between Asian languages are typically of lower quality than translations between European languages. Similarly, after eight years, Aiken [7] has updated and reevaluated the exact text used in the initial investigation, revealing a 34% enhancement as per BLEU scores, indicating that the online machine system has increased its Precision since the initial analysis. This recent research reveals that translations between English and German, Afrikaans, Portuguese, Spanish, Danish, Greek, Polish, Hungarian, Finnish, and Chinese are frequently the most precise. Furthermore, a comparison study was conducted between English-Persian and Persian-English translations from Google Translate using Keshavarz's error analysis model [11]. They found that the variation in the frequency of various English-to-Persian and Persian-to-English mistakes was not statistically significant because both had errors. Hence, the translation direction did not impact the quality of machine-generated translations.

The accuracy of Google Translate for translating a legal document from Thai to English was examined qualitatively by Vidhayasai and his co-authors [12]. According to the research, errors occur at three primary levels: lexical, syntactical, and discursive. So, they wanted to determine if Google Translate is practical or valuable. It was discovered that the translation of the document, which linked to the terms and conditions on an airline website, needed to be more accurate and easier to understand.

Yang and his co-authors proposed the adoption of appropriate weights for various words and n-grams into the traditional BLEU framework [13]. They merely added part-of-speech (POS) and n-gram length information to the conventional BLEU framework using a linear regression model to maintain language independence in the BLEU framework. Their study's experimental findings demonstrate that this augmentation produces accuracy superior to the original BLEU.

Moreover, Adly and Al-Ansary [14] evaluate Arabic machine translation using the Universal Networking Language (UNL) and Interlingua's translation methodology. The interlingua method converts text from the selected language into a representation form that may be translated into the target language. The assessment procedure included three metrics: F-mean, F1, and BLEU. The Encyclopedia of Life Support Systems (EOLSS) was used for the evaluation. Al-Ansary and his co-authors conducted studies on the impact of UNL on translation from and into Arabic.

Al-Haj et al. [15] conducted a study that examined the impact of Arabic morphological segmentation on statistical machine translation from English to Arabic. Their investigation primarily centered on phrase-based statistical machine translation. Notably, their results demonstrated a substantial influence of the segmentation approach on the performance of the statistical machine translation system, as evidenced by variations in BLEU scores between the most effective and the least effective morphological segmentation methods.

The translation aspect and the finalized texts generated by Google Translate have been the subjects of prior investigations. For example, Rasul [16] critically evaluated the Precision of translations from English to Sorani Kurdish generated by Google Translate in light of linguistic and cultural factors. For this reason, the paper has chosen to critically evaluate examples of translations produced by Google Translate using a textual analysis approach based on the concept of correlation. According to the results, Google Translate is mainly accurate in terms of morphological and syntactic accuracy. Nevertheless, improvements are required to rectify minor shortcomings found in clitics, IZAFA, and element reordering future perfect, perfect progressive tenses, third conditional sentences, and passive tenses. Regarding the semantic level, Google Translate has demonstrated accuracy while processing polysemous words and non-equivalence phrases. However, It is considered an absolute deficiency when idioms and proverbs are translated.

In addition, the study by Kaka-Khan and Taher [17] tried to assess the In-Kurdish machine translation system regarding linguistic and computational issues for the given text as uncommon words, sentences, phrases, and paragraphs. They attempted to evaluate in-Kurdish machine translation using BLEU, NIST, and METEOR-TER.

A separate study [18] used Google Translate to translate a leaflet from English into Spanish and Chinese. The findings indicated that the English-to-Spanish translation was more accurate than English-to- Chinese. Furthermore, a human translation in Spanish did not generate a prominently improved result over Google Translate.

Moreover, another study [19] used GT to translate ten medical statements in 26 languages (8 Western European, 5 Eastern European, 11 Asian, and 2 African). According to the findings, the translation accuracy was rated at 57.7%, despite two African languages obtaining the lowest translation rate (45% correct) and eight Western European languages performing the best (74%).

Swahili had the lowest correct translation rate of 10%, while Portuguese received the highest correct translation rate of 90%.

However, Bozorgian and Azadmanesh [20] suggested a new evaluation of the Google Translate study. They assessed subject-verb agreement when translating English sentences into Persian using Google Translate and human translators, and they estimated that Google Translate needs to manage subject-verb agreement with as much accuracy as human translators do.

3. Methodology:

This comparative study evaluated the quality of machine translation (the candidate) compared to human translation (the reference) by calculating the accuracy and the errors with percentages occurring in MT. Figure 2 demonstrates the primary steps performed to carry out this research. The English instances text is initially translated using Google Translate, followed by the HT (reference). After that, these Kurdish samples were then preprocessed by dividing the text into n-gram sizes, ranging from unigrams to bigrams, trigrams, and tetragrams. The NIST score, BERTScore, and Precision for the Google Translate (MT) system were computed for each four-gram size. The BLEU score can be calculated by combining the Precision of four grams.

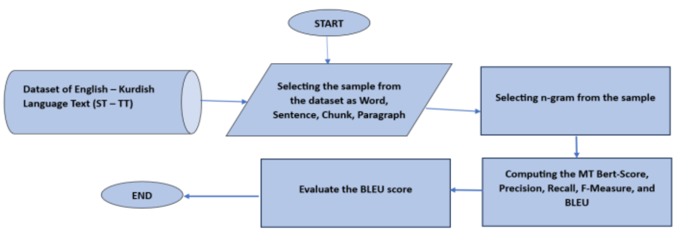


Figure 2: Flowchart of Google Translation Evaluation using BLEU and NIST Scores.

3.1 Materials

3.1.1 Online Machine Translation Tool

To find translations of English texts (ST) into the Kurdish language (TT), it seemed logical to choose the program native Kurdish speakers use the most regularly. Upon entering the search terms (free + online + translate + English + Kurdish), the Google search engine showed a list of packages. Since Google Translate is the most utilized system for instant translation, it was chosen as the first alternative in researching the several programs provided. Thus, in this experiment, we use Google Translate (Google, https://translate.google.com) English-Kurdish (Sorani) translation engine as a machine translation (candidate) to observe the difference between the Kurdish translation of English sentences compared to the sentences translated by the human translators (references).

3.1.2 Selection of Source and Target Text

In this experiment, the English language is selected as a source text, and the Kurdish language is chosen as the target text for the translation. Both languages originate from distinct language families with unique linguistic roots and cultural heritage.

Since the objective of the present research was to evaluate the efficiency of Google Translate Machine Translation (MT) in converting an English ST into Kurdish TT, we considered the quality of both the semantics and grammar in the produced STs. Furthermore, grammatical and semantic mistakes can make it challenging to understand the message when translating from the source text (ST) to the target text (TT). It is crucial to note that, given the characteristics of the two languages examined in this research, these errors may result in fidelity and comprehension issues.

Languages with much flexibility in ordering words can trigger more than one sentence for the same purpose, such as the examples in Table 1. Also, languages with many inflectional variations are one category of morphologically rich languages. They are challenging for MT as well as other applications of language technology. The main problem with widely conjugated languages is that the quantity of conjugated word forms leads to data sparsity (see Table 1), which leads to inaccurate statistical MT estimations. Most words in a corpus only appear a few times at most. As a result, there is limited coverage of the translation rules, and the translation probabilities could be better estimated. The following example shows the richness of Sorani Kurdish in terms of element reordering syntactically since one sentence can be presented in three different ways.

Table 1: Word permutation of the English sentence "Taban made the bed." The first example is used in conversational speech.

1	- Taban bed-Definite A.IZAFA-fix pstتابان جێگەکەی چاک کرد
2	BedDEF.article-Taban-Fix-IZAFA-جێگەکە تابان چاکی کرد
3	Bed-Def.-from-Taban-Fix-passiveجێگەکە لەلایەن تابانەوە چاککرا

Language-specific tools, which are not always accessible, are significantly relied upon in many works on morphology-aware techniques. Low-resource languages include a large number of morphologically rich languages. Languages with many inflectional variations are one kind of morphologically rich language. They are challenging for MT as well as other applications of language technology. The primary issue with widely conjugated languages is that the abundance of conjugated word forms causes data sparsity (see example in Table 1), which leads to inaccurate statistical MT estimations. Most words in a corpus only appear a few times at most. These phenomena lead to insufficient coverage of translation rules and inaccurate estimation of translation probabilities [21].

Regarding the present and future tenses in English and Kurdish, a significant difference is noticed in present and future markers; thus, conveying a future action is through the "here and now." This property is very similar to how English occasionally conjugates verbs in the present tense to suggest future action. Saying, "I am going to the store later, do you want to come?" is one example of what you can say. It is almost equivalent to saying, “I am going to the store later; would you like to come?” Sorani blends everything into a tense called “ڕانەبردوو” or “non-past.”

According to the information obtained from [5], language employs tenses to express the sequence of events. Tenses help answer questions like whether something happened yesterday, is happening now, or will happen in the future. To address these inquiries, one must grasp the tenses in a language. In Sorani Kurdish, a single combined present and future tenses exist. Meanwhile, it has four past tenses. Table 2 illustrates the three different English sentences for the Sorani Kurdish sentence “**پاسەکە کاتژمێر ١٠ دەگات**” for both present and future tenses.

Table 2: Word permutation of the Kurdish sentence “.پاسەکە کاتژمێر ١٠ دەگات”.

1	پاسەکە کاتژمێر ١٠ دەگات Paseke katjmer 10 degat	The bus arrives at 10 o’clock.
2		The bus is arriving at 10 o’clock.
3		The bus will arrive at 10 o’clock.
		Bus. Art. o’clock 10 (present simple, continuous, and future marker) arrives.

The most significant aspect of the English language in this context is its flexibility in accommodating the future marker, as demonstrated in the previously mentioned example. In contrast, Sorani Kurdish lacks distinct markers for the present simple, progressive, and future tenses, resulting in all sentences having a single form in Kurdish. In contrast, English has three distinct forms for these tenses—moreover, present and future tenses in Sorani Kurdish share a similar structure. When translating the provided English examples into Sorani Kurdish, all three sentences are rendered in a single form, with comprehension relying on context. The human reference and the candidate translation produced similar results, aligning with the reference translation. Ibid additionally mentioned that Sorani Kurdish features four different past forms.

3.2 Sample Dataset

To evaluate the machine translation, we first prepared a corpus consisting of the random source text of English as words, sentences, compound words, and paragraphs, as shown in Table 3. This evaluation procedure involves the comparison of machine-translated segments with a human-translated reference text, as the quality of a machine translation is considered better when it closely resembles a professional human translation [22]. The BLEU score, a widely used metric in machine translation evaluation, assesses the translation quality of each output text. This score plays a crucial role in our research, as it helps us identify the Google Translate application that delivers higher-quality English-to-Kurdish translations.

3.3 Evaluation Metrics

Because there are various correct methods for translating a given ST, multiple approaches might be taken to assess the translated scripts; this endeavor can result in a subjective assessment of some type. Additionally, when experienced translators and evaluators view a translated document a second time after seeing it the first time, they could respond differently to the analysis topic [23].

The current evaluation process involves dividing the ST into “semantic chunks” and assessing translations based on reliability and clarity. The BLEU score, NIST Score, and BERTScore are used to measure the accuracy of each output text’s translation. This result led to the discovery of the Google Translate system, which guaranteed an improved English-to-Kurdish translation. Based on the definition, a semantic chunk is an array of words that satisfies a semantic role stated in a semantic context. As a result, each section should describe who did what to whom in some way [25]. Other elements connected to “where, when, how, or why” may also correspond to chunks.

3.3.1 Bert Score

BERTScore is an automated text-generation evaluation metric. It calculates the semantic similarity rate for each symbol in the candidate-translated sentence, $\hat{x} = \hat{x}_1, \hat{x}_1$, with each symbol in the reference translated sentence $x = x_1, x_k$, optionally weighted with frequency rate in the inverse document. Figure 3 illustrates the computation [26].

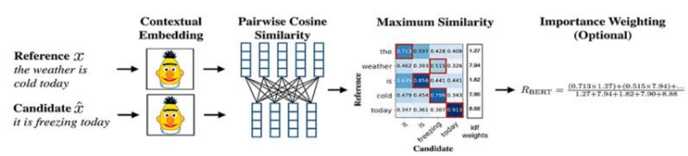


Figure 3: The RBERT recall metric computation is illustrated for the given reference and candidate text.

The Recall score is calculated by matching each token in x and a token in $\hat{x}$, and Precision is calculated by matching each token in $\hat{x}$ to a token in x. We use greedy matching to maximize the matching similarity score, where each token matches the most similar token in the other sentence. We calculate the F1-measure as a result of Precision and recall combination. The Recall, Precision, and F1 scores for the reference x and candidate $\hat{x}$ are:

$$R_{\mathrm{BERT}} = \frac{1}{|x|} \sum_{x_i \in x} \max_{\hat{x}_j \in \hat{x}} \mathbf{x}_i^\top \hat{\mathbf{x}}_j \ , \quad P_{\mathrm{BERT}} = \frac{1}{|\hat{x}|} \sum_{\hat{x}_j \in \hat{x}} \max_{x_i \in x} \mathbf{x}_i^\top \hat{\mathbf{x}}_j \ , \quad F_{\mathrm{BERT}} = 2\frac{P_{\mathrm{BERT}} \cdot R_{\mathrm{BERT}}}{P_{\mathrm{BERT}} + R_{\mathrm{BERT}}}$$

Table 3: The Sample of the prepared dataset.

POS (parts of speech)	Source Word	Human Translated	Machine Translated
Noun	Glass	شوشە، پەرداخ، جام	شووشە
	health	لەشساغی	تەندروستی
Verbs	reproach	سەرزەنشتکردن	سەرزەنشتکردن
Adjectives	carefree	بێ خەم	بێ باک
	expert	شارەزا، پسپۆڕ	شارەزا
Compound words	Assistant Lecturer	مامۆستای یاریدەدەر	یاریدەدەری وانەبێژ
Sentences	Sara runs the class very well	سارا پۆلەکە زۆر باش بەڕێوە ئەبات	سارا زۆر بە باشی پۆلەکە بەڕێوەدەبات
Paragraphs	As an expert in artificial intelligence and machine learning with more than five years of experience in diverse domains, I have improved my skills and knowledge in various areas, with a primary focus on healthcare, climate change, roads, and transportation. I have worked with large-scale datasets, such as signals, numerical data, and texts, to create a robust and accurate predictive model aiming to improve outcomes in corresponding domains.	وەک کەسێکی شارەزا لە زیرەکی دەستکرد و فێربونی ئامێریی و هەبوونی زیاتر لە ٥ ساڵ ئەزموون لە بوارە جیاجیاکان، لێهاتوویی و توانا و زانیارییەکانم باشتر کردووە بەتایبەت لە بواری چاودێری تەندروستی، گۆڕانی کەش و هەوا، ڕێگاوبان و هۆکارەکانی گواستنەوە . من کارم لەسەر چەندێن داتای جۆراوجۆر کردووە وەکو داتای هێما و دەقی ژمارەیی بە مەبەستی دروستکردنی مۆدێلێکی پێشبینیکراوی بەهێز و وورد بە مەبەستی باشترکردنی دەرەنجامەکان لە بوارە هاوبەشەکاندا.	وەک شارەزایەک لە زیرەکی دەستکرد و فێربوونی ئامێر کە زیاتر لە ٥ ساڵ ئەزموونی هەیە لە بوارە جۆراوجۆرەکاندا، لێهاتوویی و زانیارییەکانم لە بوارە جیاوازەکاندا باشتر کردووە، بە گرنگیدانێکی سەرەکی لەسەر چاودێری تەندروستی، گۆڕانی کەشوهەوا، ڕێگاوبان، و گواستنەوە. من لەگەڵ کۆمەڵە داتا گەورەکان کارم کردووە، وەک سیگناڵ، ژمارەیی، دەق، بۆ دروستکردنی مۆدێلێکی پێشبینیکردنی بەهێز و ورد بە ئامانجی باشترکردنی دەرئەنجامەکان لە دۆمەینە هاوبەشەکاندا.

Where R_{BERT} counts the number of correctly translated words compared to the machine-translated words, P_{BERT} counts the number of candidate translation words (unigrams) that occur in any reference translation by the total number of words in the candidate translation. The F-measure (F_{BERT}) of the translation is equal to the multiplication of the Precision and recall divided by the addition of the Precision and Recall [21],[27-28].

3.3.2 BLEU Score

The BLEU (Bilingual Evaluation Understudy) score is a typical evaluation score used to measure the performance of machine translation models. A BLEU implementor's main programming task is determining the similarity or semantic closeness between the machine (candidate) and the actual Human (Reference) Translation. The BLEU score calculates the Precision of accumulated lexical matches for n-grams, considering n-grams of lengths up to four [22]. This calculation is accomplished by examining the n-grams in the candidate text and comparing them to the n-grams in the reference translation, subsequently counting the number of matches. These matches are position-independent. The more matches, the better the candidate translation is [21].

The BLEU score falls within a scale of 0 to 1, with a higher score signifying a more robust correspondence between the machine translations and the reference text. A BLEU Score close to 1 means that the model's translations are almost identical to the reference translations. A score close to 0 means that the model's translations differ significantly from the reference translations [29]. BLEU is generally calculated over the complete corpus rather than individual sentences. It is essential to recognize that only some translations will achieve a score of 1 unless they match a reference translation precisely. Hence, a human translator may receive a poor score of 1, as substantial variability exists in acceptable translations [21],[30].

The formula for calculating the ultimate BLEU score, denoted as formula (2), is presented. It relies on the Brevity Penalty (BP), which involves selecting the effective reference length, represented as "r." This reference length is further detailed in formula (1) within Figure 4 [30].

3.3.3 NIST

The NIST score is an adjusted variant of BLEU that assigns varying weights to each n-gram and determines the brevity penalty based on the shortest reference. NIST scores are computed for various n-gram lengths, typically with a default value of 5 [10],[17].

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{\left(1-\frac{r}{c}\right)} & \text{if } c \leq r \end{cases}$$

Formula 1. *Brevity Penalty (BP)*

$$BLEU = BP \times \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

Formula 2. *BLEU score, where N=4 and uniform weights wn=(1/N), P is the precision*

Figure 4: BLUE Score formula with Brevity Penalty.

4. Results and Discussions

According to Aitchison [31], linguistics is the scientific study of language, and language is the way of communication. Linguistics has various scopes, and this study attempted to deal with three levels of linguistics: morphology, syntax, and semantics. These scopes are defined and investigated using natural language processing to evaluate Google Translate between human and machine translation.

Morphology studies the internal structure of words and the rules by which they are formed; morphemes are the smallest unit in a language. For example, the word (tourists) comprises three morphemes: tour+ist+s. Morphemes are free and bound morphemes [32].

Tallerman [33] declared that nouns fall into different genders or noun classes in many languages. However, the English and Sorani Kurdish languages do not have gender, and the nouns are not indicated by gender. In other words, this allows the machine to perform the translation with fewer shortcomings since both languages have similar features regarding gender.

Syntax can be defined as “sentence construction” and how words are grouped to make phrases and sentences; it also means the study of the syntactic properties of language [33]. Moreover, in some languages, morphology and syntax merge to shorten sentences and create a feature called morphosyntax, found in Sorani Kurdish. Morphosyntactic changes occur due to changes in syntactic structures along with morphological derivations or inflections, which arise in Kurdish but are not found in English. Moreover, as mentioned in [34], computational linguistics is the study of computers that simulate language, and it is patterned. Semantics is defined as the field of linguistics that links together the sound patterns and the meaning, and syntax and semantics are the ‘bread and butter” of linguistics and are a central concern in this paper, too [Ibid]. Regarding Morphology, the machine cannot specify the right word for polysemous words, such as:

Glass	Human Reference (HT) (شوشه، پەرداخ، جام)	MT(Machine Candidate)شووشه

Hence, it is possible to translate a given source text (ST) in various correct ways, and consequently, there are multiple approaches to assess the translated texts. This strategy can introduce an element of subjectivity into the evaluation process. Furthermore, professional translators and evaluators may react differently to the content of a translated text when they encounter it for the second time compared to their initial assessment [23].

During the current evaluation, it was crucial to break down the source text (ST) into “semantic chunks” and evaluate the translations according to their clarity and accuracy. A semantic chunk is described as “a sequence of words that fulfills a semantic role specified within a semantic frame.” Therefore, each chunk is expected to convey some or all of the information regarding “who did what to whom?” [25].

This study evaluated single words, compound words, sentences, and paragraphs of randomly selected samples from the testbed. The accuracy is calculated using evaluation metrics such as Precision, Recall, F-measure, BLEU Score, and NIST score, and the similarity matrix of each sample is described accordingly.

The distinctions observed between the machine-translated (MT) candidate text and the human-translated (HT) reference text in the chosen testbed samples encompassed various linguistic aspects, such as vocabulary selection, word order within compound words, the usage of singular and plural forms of nouns/pronouns, and the application of inflections, both at the sentence and paragraph levels.

4.1 Word Choice Translation Machine Evaluation

Among many words in the dataset, the random single word “bag” has been selected as a testbed for investigation, and the metric scores with similarity matrix of selected words of different linguistic types are illustrated.

4.1.1 BLEU and NIST Scores

Since words consist of one part unigram, the evaluation between the human translation and machine translation gives more accurate results. As shown in Table 4 below, the BLEU score of the selected word “bag” is 1, which means the HT (reference or Human Translation) and MT(candidate or Machine Translation) are identical. Consequently, Google Translate is excellent for translating words according to NLP metrics.

Table 4: The BLUE and NIST scores for the selected word “bag”.

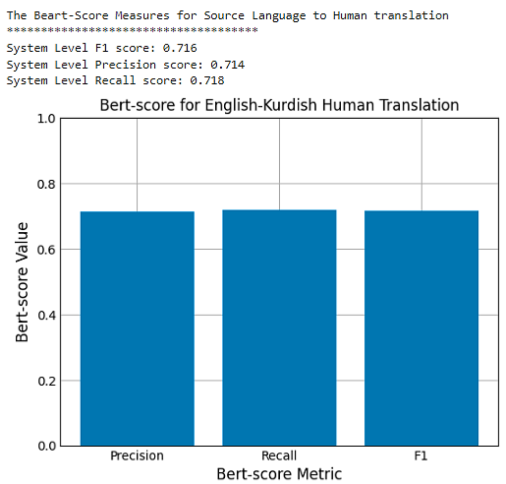
Source	Machine	Human (Reference)	BLEU SCORE	NIST SCORE
bag	جانتا	جانتا	1.00	2.17

4.1.2 Bert Score

For the selected word "bag" in the dataset, the BERTScore for the MT (candidate) and HT(reference) is illustrated in Figure 6.

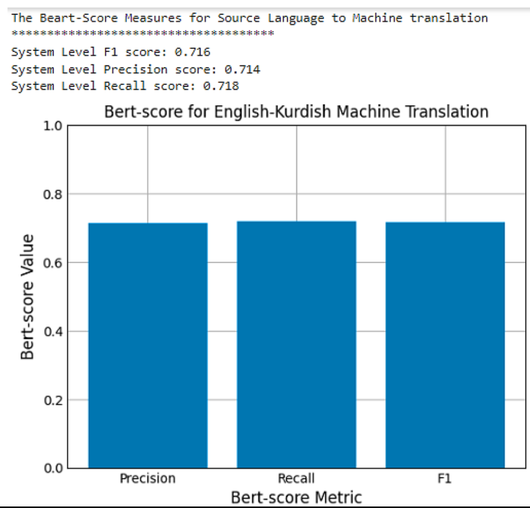
The value of the Bert Score metrics (F1, Precision, and Recall) for the source to HT (reference) is shown in Figure 5. a. Precision indicates the count word "bag," which is correctly translated by MT, and Recall suggests the number of MT words. The F-Measure is the product of both Recall and Precision. The result of the (F1, Precision, and Recall) for the source to MT (candidate) is indicated in Figure 5. b. The values are the same for MT (candidate) and HT (reference), which suggests that Google Translate gives the best result for word translation.

F1	Precision	Recall
0.716	0.714	0.718



a. Source to Human Translation BertScore.

F1	Precision	Recall
0.716	0.714	0.718

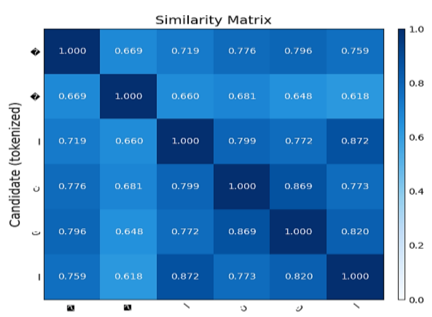


b. Source to Machine Translation BertScore.

Figure 5: BERTScore illustration for the selected word "**Bag**".

4.1.3 Similarity Matrix

The similarity matrix for the selected word "bag" from the dataset is illustrated in Figure 5. The x-axis represents the letters of the HT (reference) word, and the Y-axis represents the MT (candidate) word. Since the n-gram of the MT (candidate) and n-gram of the HT (reference) match perfectly and give the result 1, as illustrated in Figure 6, there is no difference in word-by-word translation between the human reference and machine candidate translation. The more matches, the better the candidate translation is. The Kurdish language is one of the undocumented languages; according to [35], some of the matching letters in the selected word are not correctly recognized, and despite this, the machine calculates them as identical.



a. Similarity Matrix Before Rescaling

b. Similarity Matrix after Rescaling

Figure 6: Similarity Matrix for the word "bag" between MT (candidate) and HT (Reference).

4.2 Compound Word

4.2.1 BLEU and NIST Scores

There are several sizes to measure the accuracy of compound words: unigrams, bigrams, trigrams, and tetra-grams. In our dataset, the compound word "tall order" has been selected for evaluation. Based on the example chosen, the translation from English into Sorani Kurdish by MT (candidate) and HT (reference) is different, giving the result of 0 BLEU SCORE and 0.89 NIST score, as shown in Table 5. Linguistically, it can be shown that the difference in the translation of both MT and HT is considered meaningful when they are used as single words or dictionary meaning, but for compound words, which means having more than one meaningful word joined together, or contextual meaning, the overall linguistic value of the translated compound words or textual meaning becomes vague or different from the source text. As shown in the data, the regulators of GT must be more considered with the contextual meaning to allow the users of Sorani Kurdish on GT to be more confident in their translation. Henceforth, the semantic level, along with syntactic and morphological levels, should be considered.

Source	Machine	Human References	BLEU SCORE	NIST SCORE
tall order	فەرمانی بەرز	کاری قورس	0.00	0.89

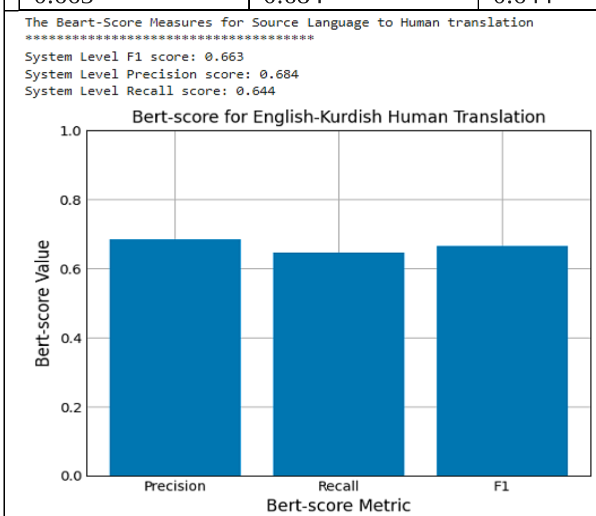
Table 5: The BLUE and NIST scores for the selected compound word "tall order".

4.2.2 Bert Score

For the selected compound word "tall order" in the dataset, the BERTScore for the MT (candidate) and HT (reference) is illustrated in Figure 7. The value of the Bert Score metrics (F1, Precision, and Recall) for the source to HT (reference) is shown in Figure 7. a, and the result of the (F1, Precision, and Recall) for the source to MT (candidate) is shown in Figure 7. b. Since the outcomes are different for machine and human translation, Google Translate gives the best result for word translation.

Source to Human Translation Bert Score

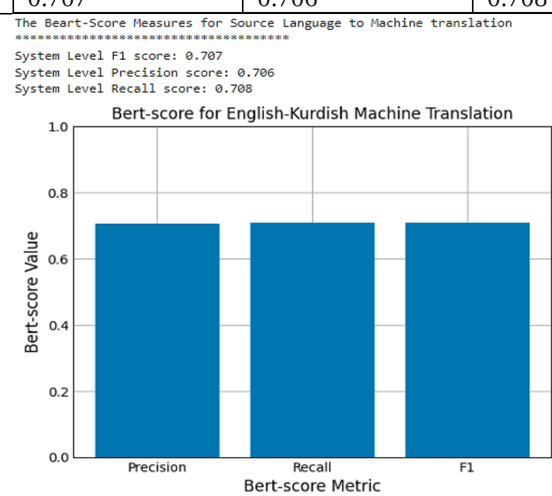
F1	Precision	Recall
0.663	0.684	0.644



a. BERTScore for English-Kurdish Human Translation

Source to Machine Translation Bert Score

F1	Precision	Recall
0.707	0.706	0.708

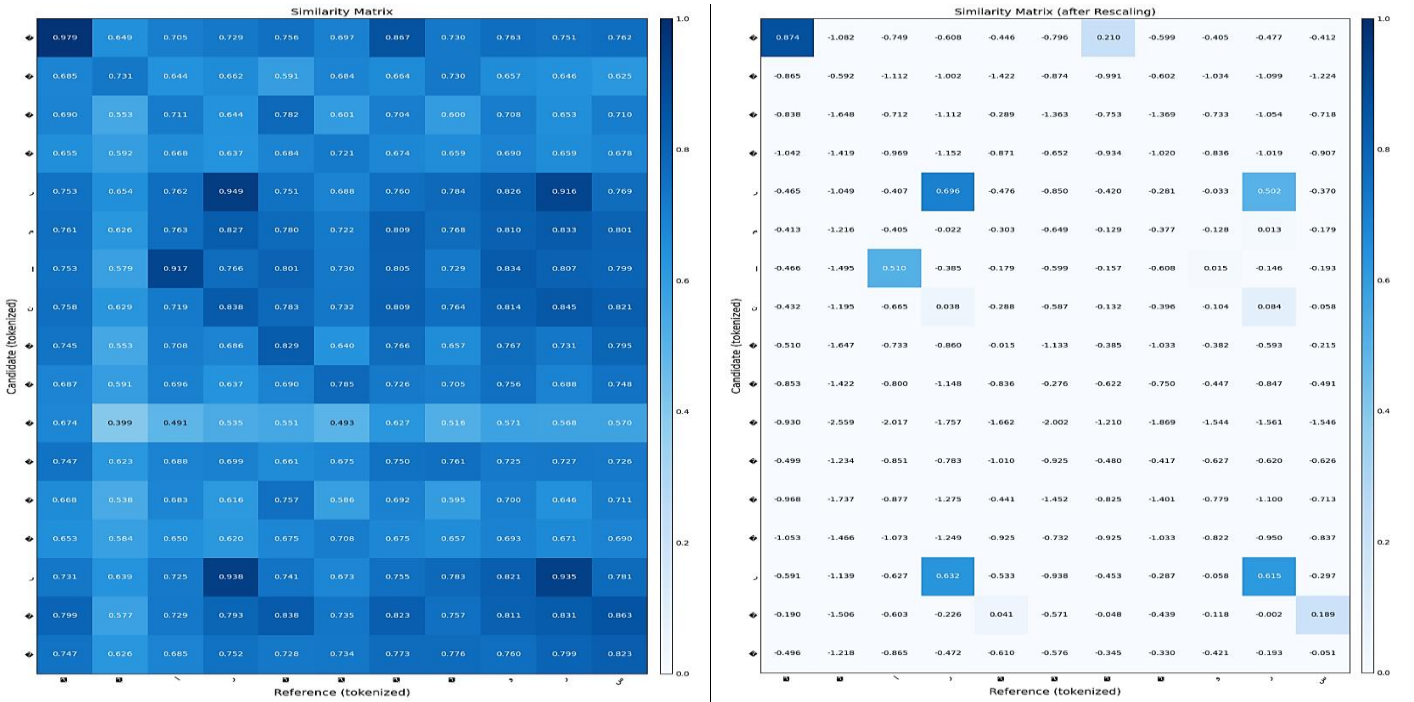


b. BERTScore for English-Kurdish Machine Translation

Figure 7: BERTScore illustration for the selected compound word "tall order".

4.2.3 Similarity Matrix

Semantically, contextual meaning should replace the literal meaning to avoid mistranslation, whether in a phrase, a sentence, or a paragraph, to provide the best sense of the translated word from the source language to the target language. In the following example, "tall order," the similarity for both MT (candidate) and HT (reference) is illustrated in Figure 8. The X-axis represents the reference translated letters, and the Y-axis represents the candidate translated letters. The highlighted cells in Figure 8. b, from the right-bottom to left-top (diagonally), show the similarity points for the HT and MT translated letters, which is done literally, and the contextual meaning is missing and causes mistranslation.



a. *Similarity Matrix Before Rescaling*

b. *Similarity Matrix after Rescaling*

Figure 8: There is no similarity between the Human (Reference) and Machine (Candidate).

4.3 Compound Word

4.3.1 BLEU and NIST Scores

IZAFA is a syntactic feature of the Kurdish language but does not exist in English. The example *"Blind confidence"* is another randomly selected compound word from the dataset for the evaluation. As shown in Table 6, the BLEU score for the MT and HT gives the result of 0.23, and the NIST score is 1.55. However, the result of the BLEU score is less than 0.5, which indicates that it could be a better translation. Nevertheless, the Google Translate MT system can recognize IZAFA when translating a text from English into Kurdish as "ی," which is transcribed as "i" or "y" in Sorani Kurdish.

Table 6: The BLUE and NIST scores for the selected compound word "Blind Confidence".

Source	Machine	Human References	BLEU SCORE	NIST SCORE
Blind Confidence	متمانەیەکی کوێرانە	متمانە ی تە واو	0.23	1.55

4.3.2 BERTScore

As IZAFA is a prominent syntactic feature of Sorani Kurdish, the system gives different values to the HT (reference) and MT (candidate). For the selected compound word "blind confidence" in the dataset, the BERTScore for the MT (candidate) and HT (reference) is illustrated in Figure 9. The value of the Bert Score metrics (F1, Precision, and Recall) for the source to HT (reference) is shown in Figure 9. a, and the result of the (F1, Precision, and Recall) for the source to MT (candidate) is shown in Figure 9. b. Due to the different meaning of 'blind' within the compound word "blind confidence," it gives an ill-formed result in both HT (reference) and MT (candidate), so the BERTScore measures for both source-to-human and source-to-machine are different.

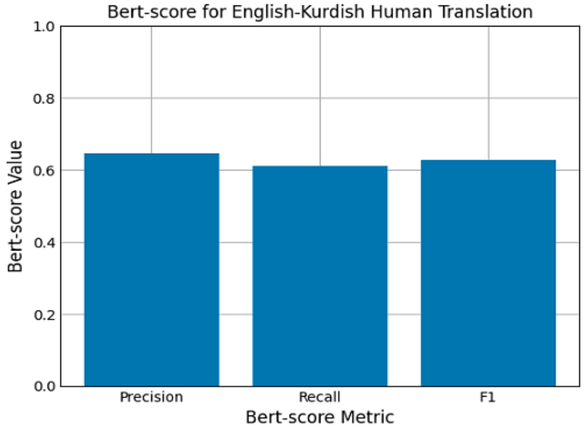
4.3.3 Similarity Matrix

The similarity for the selected compound word "Blind Confidence" is illustrated in Figure 10. The X-axis represents the reference-translated letters, and the Y-axis represents the candidate-translated letters. The highlighted cells in Figure 10. b, from the right-bottom to left-top (diagonally), show the similarity points for the HT and MT translated letters. Because meaning is the core issue of semantics and translation simultaneously, the meaning of confidence in the selected example is similar in both HT and MT; meanwhile, the meaning of blind is not delivered correctly and causes the difference between the MT and HT letters.

Source to Human Translation Bert Score

F1	Precision	Recall
0.627	0.645	0.610

```
The Beart-Score Measures for Source Language to Human translation
*************************************
System Level F1 score: 0.627
System Level Precision score: 0.645
System Level Recall score: 0.610
```

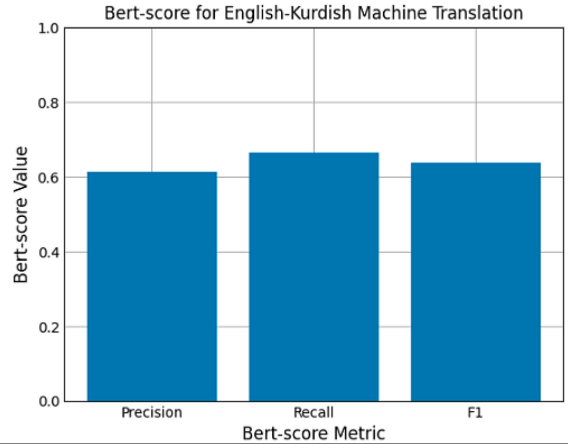


a. BERTScore for English-Kurdish Human Translation

Source to Machine Translation Bert Score

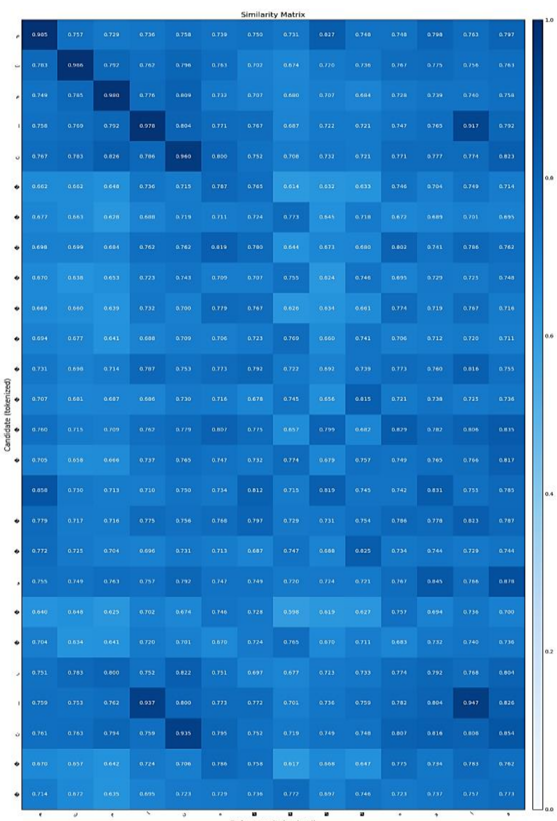
F1	Precision	Recall
0.638	0.614	0.664

```
The Beart-Score Measures for Source Language to Machine translation
*************************************
System Level F1 score: 0.638
System Level Precision score: 0.614
System Level Recall score: 0.664
```

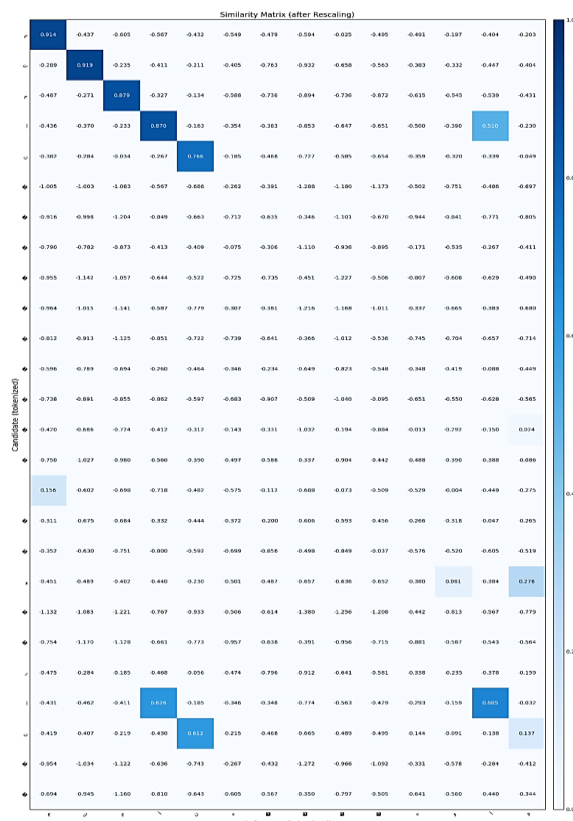


b. BERTScore for English-Kurdish Machine Translation

Figure 9: BERTScore illustration for the selected compound word "Blind Confidence".



a. Similarity Matrix Before Rescaling



b. Similarity Matrix after Rescaling

Figure 10: Similarity between the Human (Reference) and Machine (Candidate).

4.4 Sentence Translation Machine Evaluation

Regarding the sentences, some issues occur with the word order of English and Kurdish since both systems syntactically are different from each other; specifically, the syntactic structure of English and Kurdish languages are entirely different; thus, the Kurdish sentence structure consists of a subject followed by an object and a verb. However, the sentence structure of English is ordered in this way: a subject is followed by a verb and followed by either an object or a complement. Meanwhile, the perfect translation of both MT (candidate) and HT (reference) is different, and the MT preceded the object in the Kurdish language, which is satisfactory since, according to Kaka-khan and Taher [17], Kurdish is a pro-drop language, and the subject can be dropped. In this study, the sentence “He ate the cookies on the couch” was selected randomly from the NLP algorithm, and the metrics were evaluated as described in the sections below.

4.4.1 BLEU and NIST Scores

The BLEU and NIST score for the randomly selected source sentence “He ate the cookies on the couch” is shown in Table 7. The HT (reference) translated this sentence (He ate the cookies on the couch) with a sense close to a perfect translation without any oddness. However, morphologically, the word ‘cookies’ has been mistranslated by the MT (candidate) but correctly translated by the HT (reference). So, the BLEU score gives the result of 0.35, and NIST gives the result of 2.17, which is proof of the linguistic analysis.

Table 7: The BLUE and NIST scores for the selected sentence “He ate the cookies on the couch”.

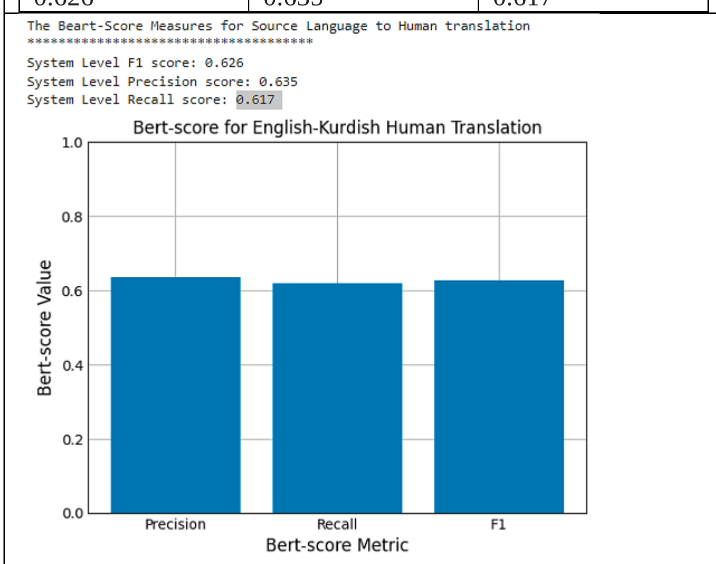
Source	Machine Translation Candidate	Human Translation References	BLEU SCORE	NIST SCORE
He ate the cookies on the couch	کولێرەی لەسەر قەنەفەکە لەسەر قەنەفەکە کولێرەکانی, خوارد خوارد	ئەو کورە پسکیتەکانی لە سەر قەنەفەکەی خوارد	0.35	2.17

4.4.2 BERTScore

For the selected sentence “He ate the cookies on the couch” in the dataset, the BERTScore for the MT (candidate) and HT (reference) is illustrated in Figure 11. The value of the BERTScore metrics (F1, Precision, and Recall) for the source to HT (reference) is shown in Figure 11. a, and the result of the (F1, Precision, and Recall) for the source to MT (candidate) is shown in Figure 11. b. The BERTScore metrics indicate the correct translated word by the MT and count out the mistranslated one, which is the word “cookies” and mistranslated by MT (candidate) as “کولێرە,” instead of the correct translated word “بسکیت” as HT (reference). The result shows that the meaning is delivered despite some MT (candidate) mistranslations.

Source to Human BertScore

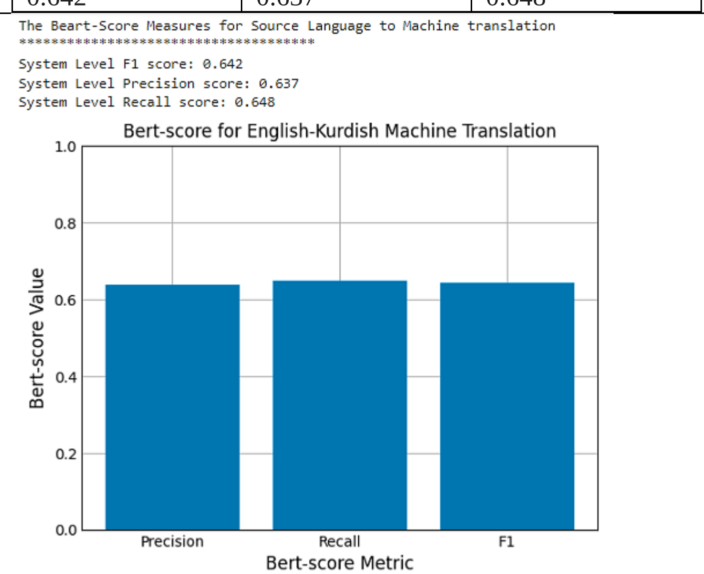
F1	Precision	Recall
0.626	0.635	0.617



a. BERTScore for English-Kurdish Human Translation

Source to Machine BertScore

F1	Precision	Recall
0.642	0.637	0.648

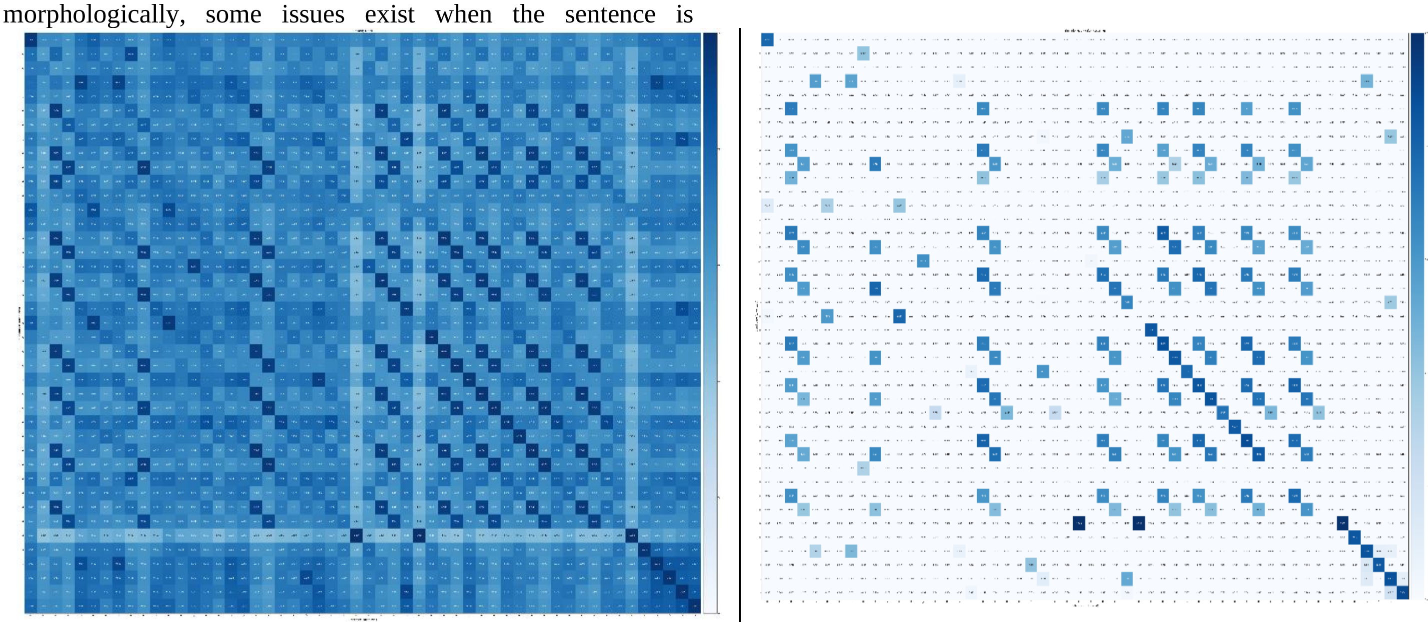


b. BERTScore for English-Kurdish Machine Translation

Figure 11: BERTScore illustration for the selected sentence: “He ate the cookies on the couch”.

4.4.3 Similarity Matrix

The similarity matrix of both MT (candidate) and HT (reference) for the selected sentence "He ate the cookies on the couch" is illustrated in Figure 12. The X-axis represents the HT (reference) translated letters, and the Y-axis represents the MT (candidate) translated letters. The highlighted cells in Figure 12. b, from the right-bottom to left-top (diagonally), show the similarity points for the HT and MT translated letters. It shows that morphologically, some issues exist when the sentence is translated by MT (candidate), not HT (Reference or Human Translation), such as adding or clipping some words. For example, when a human translates the source text into Kurdish, it is precise and complete. However, when Google Translate translates the source text, it adds or clips words. For example, the number of words included in the translation by MT increased because extra words were added to the translation to convey the message to Sorani Kurdish speakers, which can be considered a shortcoming and adds unnecessary or irrelevant information.



a. Similarity Matrix Before Rescaling

b. Similarity Matrix after Rescaling

Figure 12: Similarity between the Human (Reference) and Machine (Candidate).

4.5 Sentence Translation Machine Evaluation

4.5.1 BLEU and NIST Score

Reordering is allowed and satisfactory in the Kurdish language, unlike in English, which is strict in ordering [36]. The Kurdish translated version of the selected English sentence "**My friend translated a text from English into Kurdish**" is an obvious example and preceding the word "text" in Kurdish with the word English is overtly seen in the target language. The BLEU score gives the result of 0.6 and NIST 4.77, as demonstrated in Table 8. It shows that the MT (candidate) offers a good translation close to the HT (reference), with slight differences.

Table 8: The BLUE and NIST scores for the selected sentence "My friend translated a text from English into Kurdish".

Source	Machine Translation	Human References	BLEU SCORE	NIST SCORE
My friend translated a text from English into Kurdish.	هاوڕێکەم دەقێکی لە ئینگلیزییەوە وەرگێڕا بۆ کوردی	هاوڕێکەم دەقێکی وەرگێرا لە ئینگلیزیەوە بۆ سەر زمانی کوردی	0.65	4.77

4.5.2 Bert Score

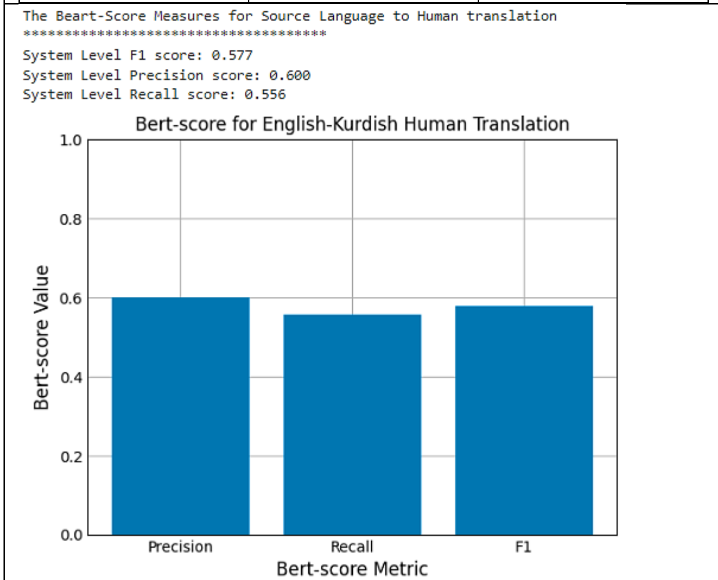
The BERTScore metrics for the randomly selected sentence in the dataset "My friend translated a text from English into Kurdish" are illustrated in Figure 13. The value of the BERTScore metrics (F1, Precision, and Recall) for the source to HT (reference) is shown in Figure 13. a, and the result of the (F1, Precision, and Recall) for the source to MT (candidate) is shown in Figure 13. b. In BERTScore, the syntactic evaluation of sentences can be translated. However, they could be more perfectly ordered due to the syntactic differences between the two languages, such as the freedom to reorder sentences in Kurdish but the stricter rules in English.

4.5.3 Similarity Matrix

Figure 14 illustrates the similarity matrix of both MT (candidate) and HT (reference) for the selected sentence, "**My friend translated a text from English into Kurdish**." The HT (reference) translated letters are represented by the X-axis (which stands for translated words by the reference), and the MT (candidate) translated letters are represented by the Y-axis (which stands for translated words by the machine). The highlighted cells in Figure 14. b, from the right-bottom to left-top (diagonally), show similar points for the HT and MT translated letters. However, there are differences in word order in MT and HT, and there is a satisfactory similarity between MT (candidate) and HT(reference), as visualized in Figure 14. b after rescaling.

Source to Human Bert Score

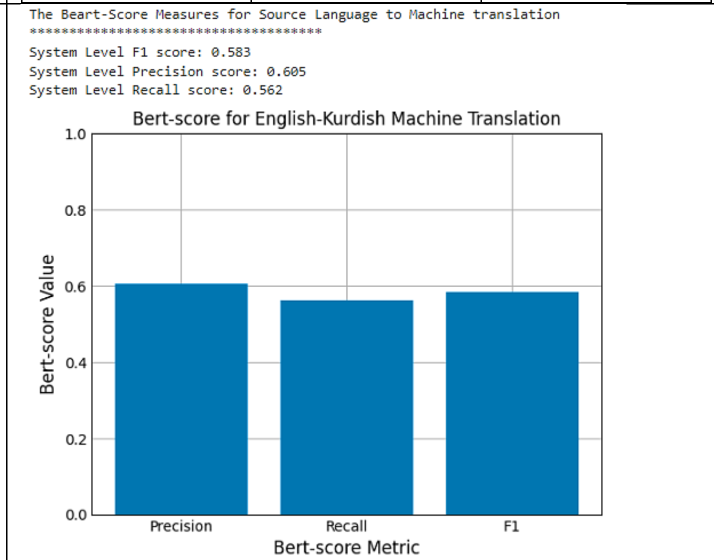
F1	Precision	Recall
0.577	0.600	0.556



a. BERTScore for English-Kurdish Human Translation

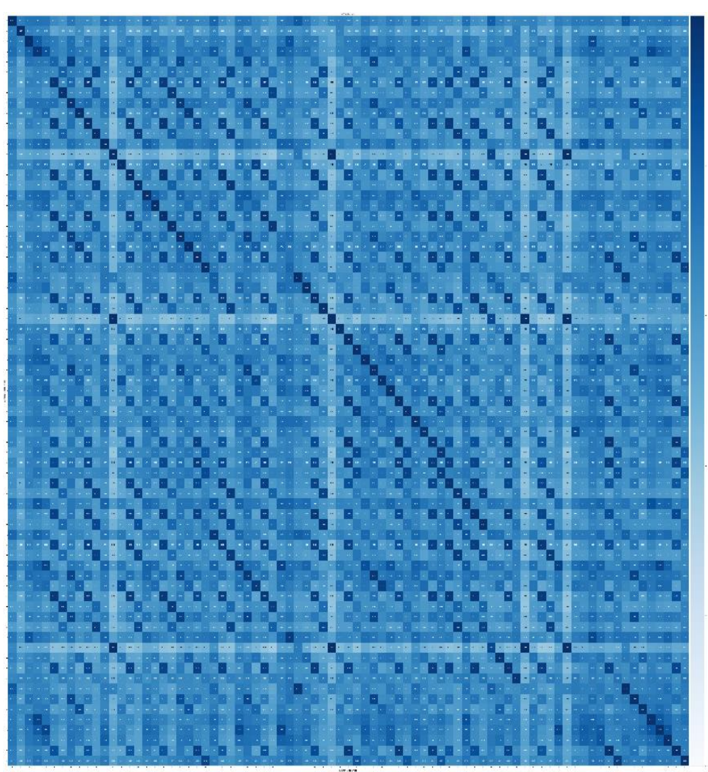
Source to Machine Bert Score

F1	Precision	Recall
0.583	0.605	0.562

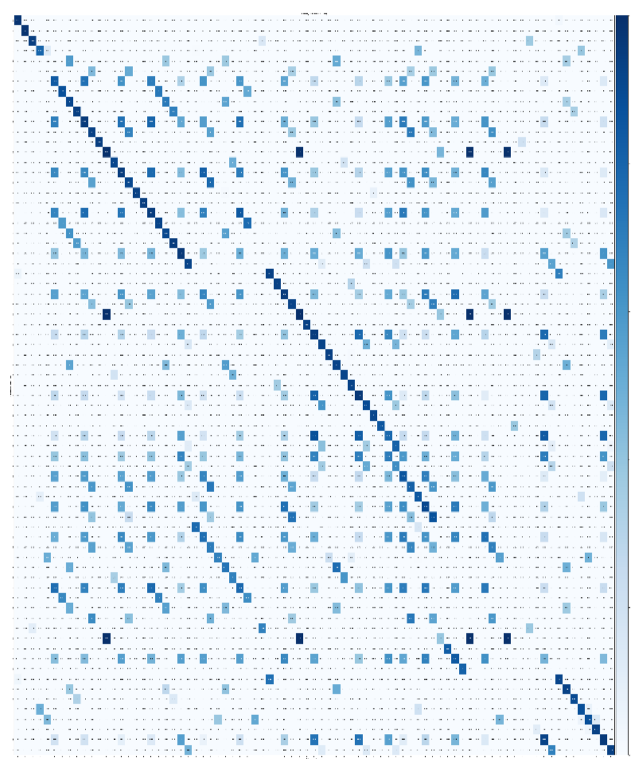


b. BERTScore for English-Kurdish Machine Translation

Figure 13: BERTScore illustration for the selected sentence “My friend translated a text from English into Kurdish”.



a. Similarity Matrix Before Rescaling



b. Similarity Matrix after Rescaling

Figure 14: The similarity between the Human (Reference) and Machine (Candidate).

Table 9: The BLUE and NIST scores for the selected paragraph.

Source Text	Machine Translation (Candidate)	Human Translation (References)	BLEU SCORE	NIST SCORE
As an expert in Artificial Intelligence and Machine Learning with more than five years of experience in diverse domains, I have improved my skills and knowledge in various areas, with a primary focus on healthcare, climate change, road, and transportation. I have worked with large-scale datasets, such as signals, numerals, and texts, to create a robust and accurate predictive model aiming to improve the outcomes in corresponding domains.	وەک شارەزایەک لە زیرەکی دەستکرد و فێربوونی ئامێر کە زیاتر لە ٥ ساڵ ئەزموونی هەیە لە بوارە جۆراوجۆرەکاندا، لێهاتوویی و زانیارییەکانم لە بوارە جیاوازەکاندا باشتر کردووە، بە گرنگیدانێکی سەرەکی لەسەر چاودێری تەندروستی، گۆڕانی کەشوهەوا، ڕێگاوبان، و گواستنەوە. من لەگەڵ کۆمەڵە داتا گەورەکان کارم کردووە، وەک سیگناڵ، ژمارەیی، دەق، بۆ دروستکردنی مۆدێلێکی پێشبینیکردنی بەهێز و ورد بە ئامانجی باشترکردنی دەرئەنجامەکان لە دۆمەینە هاوبەشەکاندا.	وەک کەسێکی شارەزا لە زیرەکی دەستکرد و فێربونی ئامێریی و هەبوونی زیاتر لە ٥ ساڵ ئەزموون لە بوارە جیاجیاکان، لێهاتوویی و توانا و زانیارییەکانم باشتر کردووە بەتایبەت لە بواری چاودێری تەندروستی، گۆڕانی کەش و هەوا، ڕێگاوبان و هۆکارەکانی گواستنەوە . من کارم لەسەر چەندێن داتای جۆراوجۆر کردووە وەکو داتای هێما و دەقی ژمارەیی بە مەبەستی دروستکردنی مۆدێلێکی پێشبینیکراوی بەهێز و وورد بە مەبەستی باشترکردنی دەرەنجامەکان لە بوارە هاوبەشەکاندا.	0.75	7.05

4.6 Paragraph Example

The Linguistic evaluation for the randomly selected paragraph, "**As an expert in Artificial Intelligence and Machine Learning with more than five years of experience in diverse domains, I have improved my skills and knowledge in various areas, with a primary focus on healthcare, climate change, road, and transportation. I have worked with large-scale datasets, such as signals, numerical, and texts, to create a robust and accurate predictive model aiming to improve the outcomes in corresponding domains**," is explained in the section below as well as its NLP metrics for MT (candidate) and HT (reference).

4.6.1 BLEU and NIST Score

The BLEU score and NIST score for the selected paragraph are presented in Table 9. The English language has less morphological reordering, and the order is considered fixed due to linguistic features of the English language on the syntactic, morphological, and semantic levels. The BLUE score gives a satisfactory result of 0.75 and NIST of 7.05. Hence, the translated word ordering and some MT meanings differ from the HT (reference). This result indicates that Google Translate follows the embedded syntactic features that support the system through NLP programming and languages.

Furthermore, as [21] stated, languages with intricate morphological structures pose challenges in machine translation (MT), mainly when translating from a language with more straightforward morphology, like English, to a language with more complex morphology, such as Sorani Kurdish. In cases where the source language lacks morphological distinctions,

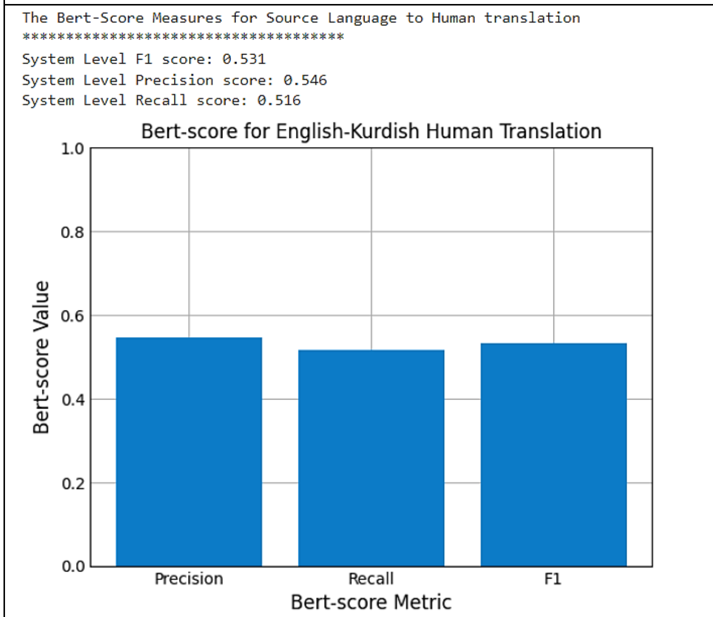
these distinctions must be created in the target language. Many approaches considering morphology depend significantly on language-specific tools, which may not always be accessible. Accordingly, the agglutinative feature of the Kurdish language is considered evident because patterns may be combined freely, whereas, in English, all patterns must be separated. For example, an agglutinated word such as those two words which are translated by the MT (candidate) زانیارییەکانم that transcribed as knowledge.DEF. PL. I first ps singular, translated as two separate patterns (My Knowledge). However, another word, which is گرنگیدانێکی is also transcribed as care. Progressive.IZAFA.give. Progressive IZAFA. (Agglutinated in Kurdish with affixes that carry one meaning, caring and similar to the English word in meaning. In addition, in syntax, words, phrases, and clauses are strung together to create coherent sentences. So, the agglutinative feature of Kurdish (Sorani dialect) also provides coherence.

4.6.2 BERTScore

The BERTScore metrics for the randomly selected previous paragraph are illustrated in Figure 15. The value of the BERTScore metrics (F1, Precision, and Recall) for the source to HT (reference) is shown in Figure 15. a, and the result of the (F1, Precision, and Recall) for the source to MT (candidate) is demonstrated in Figure 15. b. As the BERTScore Precision and Recall indicate the correct translated word from the MT (candidate) along with the length of the machine-translated words, there is a slight difference between the results, which is satisfactory. It shows that the MT (candidate), Google Translate, gives a good translation result with reasonable meanings near the HT (reference)In translating, pragmatic meaning is essential. Although occasionally the vocabulary of the source language may not be equivalent to the target language, resulting in ambiguity, meaning is therefore disclosed in context. As a result, the MT (candidate) translated the word "domains" in the given paragraph above in the same way and borrowed it from the English language while writing it in Kurdish script; however, the Kurdish word for "domains" is "(بوارەکان)," which is used by the HT (reference) and is precisely the same in meaning and a perfect word choice. When a word cannot be translated from ST into TT, the borrowing procedure is permitted; however, it is not permitted when the equivalent can be translated from ST into TT. For example, in the paragraph above, the word "signal" is translated by MT (candidate) as "(هێما)" and it is also translated as "(هێما)" by HT (reference).

Source to Human Translation BERTScore

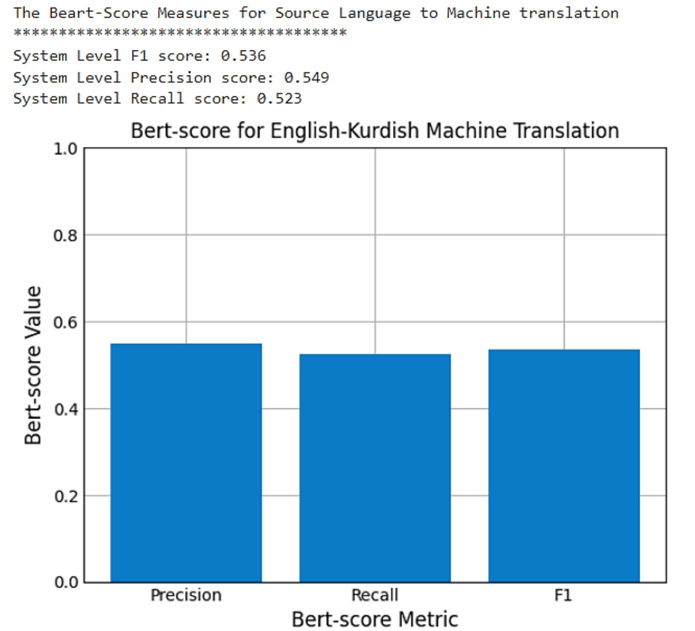
F1	Precision	Recall
0.531	0.546	0.516



a. BERTScore for English-Kurdish Human Translation

Source to Machine Translation BERTScore

F1	Precision	Recall
0.536	0.549	0.523



b. BERTScore for English-Kurdish Machine Translation

Figure 15: BERTScore illustration for the selected paragraph.

Additionally, NLP is combined with human translation, which is always the best and most valuable method that cannot be replaced by machine translation but can be used instead to speed up the translation process. To highlight the difference between human and machine translation, NLP metrics are used to enhance the quality of the evaluation done linguistically and technologically.

As a result, the evaluation was conducted from the standpoint of language analysis, which supports that the entire assessment is current but needs to be flawless. Linguistic analysis is a proper measurement because "linguistic analysis" can relate to many subjects. Branches of linguistic analysis like syntax, semantics, and morphology correspond to phenomena in human linguistic structures that substantially impact the dataset.

Conclusion

Consequently, launching Kurdish (Sorani dialect) in Google Translate is an important endeavor that benefits thousands of Sorani Kurdish speakers. This study analyzed many examples translated from English to Kurdish (Sorani Dialect). Google Translate is accurate regarding single words, compound words, and simple sentences. In other words, it can be trusted morphologically and syntactically for less complicated materials. However, semantically complex words, sentences, or paragraphs cannot be translated correctly due to flaws and deficiencies.

Accordingly, this platform needs to be enhanced to ensure that speakers of Sorani Kurdish can take advantage of Google Translate. Several syntactic characteristics, such as pronominal clitics and morphosyntactic features, could lead to a mistranslation in Sorani Kurdish.

The Kurdish language, however, could be improved if English and Kurdish experts worked together to incorporate all new situations of Kurdish (Sorani Dialect), not only semantically for polysemous words but also morphologically. The Kurdish language's morphological, syntactic, and semantic features are likely to influence some instances of translation.

Another problem is the significant cultural difference between Kurdish and English, making it difficult for machine translation software to recognize some words. The shortcomings of this system stem from the need for prior reasonable machine translation systems for the Sorani dialect of the Kurdish language, so this is a new step, and critical comments are appropriate. Overall, Google Translate reviewers should consider that this is the first attempt, so it has shortcomings and deficiencies. However, any improvement can make it a more reliable machine translation system. It can be used for translation since it is quicker than human translation, despite human translation being more accurate and having fewer errors. Nevertheless, technological advancements may help this machine translation method get closer to being faultless.

Furthermore, a very positive aspect of Google Translate as a machine translation system is that it is constantly being improved. Many errors or morphological, syntactic, and semantic shortcomings have gradually been enhanced through testing samples on GT. Therefore, linguists, experts in both English and

Kurdish, web designers, and qualified translators are helping to develop this MTS for the Sorani dialect of the Kurdish language.

In conclusion, Google Translate is used as a machine translation system. It performs admirably, but compared to human translation, the system is less accurate due to several factors, such as intellectual issues and the system's flaws in providing equivalents in Kurdish (Sorani Dialect). However, a human translation made by someone with a strong command of both languages can be trusted to be accurate and reliable. While machine translation is faster than human, human translation can still perform better, and it will only partially replace human translation since it requires human revision, even though machine translation is getting closer to producing perfect translations.

Finally, it should be noted that even though Google Translate is a successful machine translation used worldwide, Sorani Kurdish speakers can also use it for translation purposes. While it is suitable as a machine translation, it is always evident that a machine translation cannot replace a human translation. However, machine translations are quicker and more straightforward than human translations despite being less accurate and understandable [1]. Machine translation systems strike a compromise between efficacy and efficiency. Additionally, Google Translate is favored because it is a quick and efficient translation tool.

Considering the deficiencies mentioned in the study, further research on these three linguistic levels and other relevant fields is recommended and necessary to improve the accuracy of Sorani Kurdish on GT.

This paper is compared to a previous study by K.M. Kaka Khan and F.J. Taher [17], entitled "Evaluation of inKurdish Machine Translation System." The following highlights the differences between the two studies: Our current study utilizes Google Translate, whereas the prior study employed an application called inKurdish (MTS). Both studies yield remarkable results from their language evaluations of machine translation systems. Both authors curated their datasets and employed assessment instruments and measures to ensure reliable outcomes. However, compared to the other study on MTS in Kurdish, this one focusing on Google Translate garners significantly higher approval for use.

Authors contribution

All authors contributed equally to this project, from the research's implementation and design to result analysis and manuscript writing. This interdisciplinary research investigates two interwoven fields: language and technology. Sorani Kurdish has received scant attention in computational linguistics. Hence, this work aims to assess and evaluate sets of English to Sorani Kurdish translations to establish whether human translations differ from machine translations. To assess machine translation, we first created a corpus of random English original content in words, phrases, complex words, and paragraphs. Each output text's translation quality would be calculated using the NLP metrics BLEU score, NIST score, and BERTScore. In terms of morphological, semantic, and syntactic levels, a descriptive quantitative and qualitative examination of machine translation was performed, focusing on the researchers' native language intuition (Sorani Kurdish) and their linguistic background of English and Kurdish.

Funding

No external funding has supported this research.

Conflict of interests

None

Table 10: Comparing this study with previous similar studies for English-Kurdish Language Translation.

Researcher	Machine Translation Tool	Methods	Used tool	Result
[17]	inKurdish	BLEU, NIST, METEOR	Asiya toolkit	• Lack of rich corpus, • Lack of using NLP algorithm to return an appropriate result
Proposed Research	Google Translate	BLEU, NIST, BERTScore	GoogleColab witih python Language	• The rich corpus of the dataset • Using the NLP algorithm to return the best result

References

[1] Howell, N., 2019 'Applying Linguistics: Analyzing the Differences Between Human and Machine Translation of Selected Texts'. Syracuse University Honors Program Capstone Projects. 940. https://surface.syr.edu/honors_capstone/940

[2] Hassani, H. and Medjedovic, D., 2016 'Automatic Kurdish Dialects Identification', pp. 61–78. Available at: https://doi.org/10.5121/csit.2016.60307.

[3] Huang, F. and Tarp, S., 2021 'Dictionaries integrated into English learning apps: Critical comments and suggestions for improvement', *Lexikos*, 31(1), pp. 68–92. Available at: https://doi.org/10.5788/31-1-1626 .

[4] KHALID, H. S., 2019. Zimanê Kurdî,malbat û zarên wî (Kurdish language, its family, and dialects. Kurdiname, (1), 3-16. Retrieved from https://dergipark.org.tr/en/pub/kurdiname/issue/50233/637080

[5] Kim, D.H., 2010, A Basic Guide to Kurdish Grammar. Erbil: https://kurdishcentral.org/tense/?fbclid=IwAR3PagjQK21cmK3iUD3ZLIrrnA6aFu9Ubn7OR1aWutUsYcMxMR7Qz97MksU. Available at: https://kurdishcentral.org/tense/?fbclid=IwAR3PagjQK21cmK3iUD3ZLIrrnA6aFu9Ubn7OR1aWutUsYcMxMR7Qz97MksU.

[6] Hamood, M.I. and Al-bagoa, M.A.K., 2021 'The Translation of Adverbial Beta Clause Functions by Using Google Translate', 11(2), pp. 118–125. Available at: https://doi.org/10.5539/ijel.v11n2p118.

[7] Aiken, M., 2019 'An Updated Evaluation of Google Translate Accuracy,' *Studies in Linguistics and Literature*, 3(3), p. p253. Available at: https://doi.org/10.22158/sll.v3n3p253.

[8] Sabtan, Y.M.N. *et al.*, 2021 'An Evaluation of the Accuracy of the Machine Translation Systems of Social Media Language: Google's Arabic-English Translation as an Example', *International Journal of Advanced Computer Science and Applications*, 12(7), pp. 406–415. Available at: https://doi.org/10.14569/IJACSA.2021.0120746.

[9] S., L., M., T. and N., M., 2014 'Evaluating Arabic to English Machine Translation', *International Journal of Advanced Computer Science and Applications*, 5(11). Available at: https://doi.org/10.14569/ijacsa.2014.051112.

[10] Dorr, B., Snover, M. and Madnami, N., 2010 'Part 5: Machine Translation Evaluation', *Handbook of Natural Language Processing and Machine Translation*, (1), pp. 801–894.

[11] Ghasemi, H. and Hashemian, M., 2016. 'A Comparative Study of Google Translate Translations: An Error Analysis of English-to-Persian and Persian-to-English Translations,' *English Language Teaching*, 9(3), p. 13. Available at: https://doi.org/10.5539/elt.v9n3p13.

[12] Vidhayasai, T., Keyuravong, S. and Bunsom, T., 2015. 'Investigating the Use of Google Translate in "Terms and Conditions" in an Airline's Official Website: Errors and Implications', *Pasaa*, 49(January-June), pp. 137–169.

[13] Yang, M. *et al.*, 2008. 'Extending BLEU Evaluation Method with Linguistic Weight', *The 9th International Conference for Young Computer Scientists*, pp. 1683–1688. Available at: https://doi.org/doi: 10.1109/ICYCS.2008.362.

[14] Alansary, S., 2009. 'The Universal Networking Language in Action in English-Arabic Machine Translation,' *Proceedings of 9th Egyptian Society of Language Engineering Conference on Language Engineering, (ESOLEC 2009)*, (Figure 1), pp. 23–24.

[15] Al-haj, H. and Lavie, A., 2006. 'The Impact of Arabic Morphological Segmentation on Broad-coverage English-to-Arabic Statistical Machine Translation.'

[16] Rasul, S.H., 2022. 'A Critical Evaluation of Google Translation: The Case of English-Sorani Kurdish Translation', *Koya University Journal of Humanities and Social Sciences*, 5(1), pp. 114–124. Available at: https://doi.org/10.14500/kujhss.v5n1y2022.pp114-124.

[17] Kaka-Khan, K. and Taher, F., 2017. 'Evaluation of inkurdish Machine Translation System', *Journal of University of Human Development*, 3(2), pp. 862–868. Available at: https://doi.org/10.21928/juhd.20170610.39.

[18] Chen, X. *et al.*, 2016. 'Evaluating the Accuracy of Google Translate for Diabetes Education Material Corresponding Author ':, 1(1), pp. 1–11. Available at: https://doi.org/10.2196/diabetes.5848.

[19] Patil, S. and Davies, P., 2014. 'Use of Google Translate in medical communication: Evaluation of accuracy', *BMJ (Online)*, 349(December), pp. 3–5. Available at: https://doi.org/10.1136/bmj.g7392.

[20] Bozorgian, M. and Azadmanesh, N., 2015. 'A survey on the subject-verb agreement in Google Machine Translation', *International Journal of Research Studies in Educational Technology*, 4(1), pp. 51–62. Available at: https://doi.org/10.5861/ijrset.2015.945.

[21] Isabirye, M. *et al.*, 2012. ' Machine Translation and the Evaluation of Its Quality', *Intech*, p. 13. Available at: http://dx.doi.org/10.1039/C7RA00172J%0Ahttps://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics%0Ahttp://dx.doi.org/10.1016/j.colsurfa.2011.12.014.

[22] Papineni, K. *et al.*, 2002. 'BLEU: a Method for Automatic Evaluation of Machine Translation', (July), pp. 311–318.

[24] 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(05), pp. 524–548. Available at: https://doi.org/10.4236/ojml.2020.105030.

[25] White, J. and Forner, M., 2001. 'Predicting MT fidelity from noun-compound handling', *Workshop on MT Evaluation* [Preprint].

[26] Zhang, T. *et al.*, 2020. 'BERTScore: Evaluating Text Generation With Bert', *8th International Conference on Learning Representations, ICLR 2020*, pp. 1–43.

[27] Mohammed, Y. *et al.*, 2021. 'Air Temperature Prediction Using Different Datamining Approaches In Sulaymaniyah City In Iraq Air Temperature Prediction Using Different Datamining Approaches in Sulaymaniyah City in Iraq', 3(2). Available at: https://doi.org/10.24271/psr.21.

[28] Daneshfar, F., 2024. 'Enhancing Low-Resource Sentiment Analysis: A Transfer Learning Approach', 6(2), pp. 265–274. Available at: https://doi.org/10.24271/PSR.2024.440793.1484.

[29] Snover, M. *et al.*, 2006. 'A Study of Translation Edit Rate with Targeted Human Annotation', (August), pp. 223–231.

[30] Tripathi, S. and Kansal, V., 2017. 'Using Linguistic Knowledge for Machine Translation Evaluation with Hindi as a Target Language', *Computacion y Sistemas*, 21(4), pp. 717–724. Available at: https://doi.org/10.13053/CyS-21-4-2869.

[31] Aitchison, J., 2003. *Linguistics*. London: McGraw-Hill Companies.

[32] Said Ayatollah Razmjoo, 2004. *Fundamental Concepts IN Linguistics*. Tehran: Rahnama Press.

[33] Tallerman, M. , 2011. *Understanding Syntax*. 3rd edition. London: Britain published by Hodder Education.

[34] Hogg, R. and Denison, D., 2006. *A history of the English language*, *A History of the English Language*. Available at: https://doi.org/10.1017/CBO9780511791154.

[35] Fattah, M.M., 1997. *A Generative Grammar of Kurdish*. Amsterdam. Available at: https://dare.uva.nl/search?identifier=5344da45-625a-49c2-81b6-c0cbd4da9a6c.

[36] Unam, P. and Cisneros, L.O., 2018. 'VERB-GOAL (VG) WORD ORDER IN KURDISH AND NEO-ARAMAIC: TYPOLOGICAL AND AREAL CON- SIDERATIONS', p. 2018. Available at: https://www.dgcs.unam.mx/ProyectoUNAM/imagenes/180216.pdf.