JUHD / Vol. 3, No. 2, June 2017: pp862-868
DOI:10.21928/juhd.20170610.39, e-ISSN: 2411-7757, p-ISSN 2411-7765

Evaluation of inkurdish Machine Translation System*

Kanaan M.Kaka-Khan
Department Of Computer Science
University Of Human Development
kanaan.mikael@uhd.edu.iq

Fatima Jalal Taher
Department Of Translation
University Of Sulaimaniya
fatima.taher@univsul.edu.iq

Abstract

Lack of having a perfect machine translation for Kurdish language is a huge gap in Kurdish Language processing (KNLP). inkurdish is a first machine translation system for Kurdish language which is capable of translating English into Kurdish sentences. Building "inkurdish" machine translation system was a great point regarding Kurdish language processing, but like any other translation system has strengths as well as many shortcomings and issues. This paper tries to evaluate inkurdish machine translation system according to both linguistics and computational issues. It might help any other researchers interested in doing research in this field. It attempts to evaluate inKurdish from different perspectives, such as, giving un common words, sentences, phrases and paragraphs in this machine to check whether it provides the correct translation or not. A general evaluation can be done after getting a valid sample with their translations from the machine and compared to the meanings of the words outside the machine.

Keywords- NLP, Machine Translation (MT), Kurdish, Asiya Toolkit, inkurdish translator, BLEU, NIST, METEOR.

I. 1. INTRODUCTION

A. *Kurdsih Language*

Kurdish language as stated by Kurdish Academy of Languages belongs to the Indo-European family of languages. The three most widely spoken dialects of Kurdish are Central Kurdish, Kurmanji saru and Kurmanji khwaru by (Ameen, 2017), The Central Kurdish dialect uses Arabic script while the Kurmanji saru Kurdish dialect is written in Latin script. In addition, Some features of Kurdish discussed in the following definitions by different authors and linguists: firstly, Kurdish is described as one of the agglutinated languages as Mahwi (2011:13) argues that if any word of a sentence consists of two or more morphemes then this is considered as agglutination. Secondly, the basic word order of Kurdish is described by Faraj (2009:54) as Kurdish word order is SOV [subject+ object+ verb]. These three elements are in a fixed order, Sub NP, Object NP and Verb.

Thirdly, Mahwi (2011:13) declares that there is a group of languages called pro-drop languages whose subjects can be null so Kurdish has the basic structure of SOV but S can be optional. Qader (2004:56) confirms that these pro-drop languages' tense and person are lexical elements.

Finally, this feature is obviously seen in the MT analysis of some languages as Cook (1988:38ff) declares that although in Universal Grammar (UG)1 there is the parametric concern about all languages in the world, there are some languages which share some features and some others do not, such as pro-drop feature which Kurdish and Spanish both have it. Kurdish has seen less development in the field of computational linguistics, as Wahab (2015:13) states that 'computational linguistics

*** The paper has been presented in the 4th International Scientific Conference of University of Human Development, April, 2017.**

[1] UG: it is the abbreviation of Universal Grammar of Chomsky.

main job is to be a source to produce other computational programs to computerize the words and contexts in the Natural languages'.

B. Background

Machine translation is one of the most important applications of Natural Language Processing (NLP) which brought researchers attention recently. Majority of world's languages being worked on in the field of computational linguistics generally and machine translation especially, but unfortunately Kurdish language processing being the least among the world languages to be worked on, individual works have been done here and there but with no any noticeable result, majority of those attempts limited to dictionary designing for word translation up to phrase translation, recently a new attempts for Central Kurdish born which led to build a system for translating English into Kurdish texts and vice versa, this system is a web based project under the name "inkurdish". In this paper we evaluated this online machine translation system for the sake of two reasons: firstly, this work might help inkurdish designers to improve their system in coming versions or modifications, second we hope this work is capable of becoming a building block for anyone tries to design Kurdish Machine Translator later on.

MT is one of the recent research areas as (Zaretskaya, et al.,2016) state that '(Even though Machine Translation (MT) is one of the most advanced and elaborate research fields within Translation Technology, the quality of MT output has always been a great concern, and MT evaluation is a popular research topic'. However, there are some challenges the users of MT might face such as Lavie, A. (2010:3) States that 'MT Evaluation is a challenging and active research area of its own Merit'. This paper also attempts to show some drawbacks of MT such as Zaretskaya, et al.,(2016) notice that even though there is a common opinion that 'MT can be used only to get the 'gist' of the text, the development of technologies is moving forward and this idea is becoming more and more questionable'. However, in order to achieve higher quality it is necessary to be able to evaluate it.

There are several events that related to the development of MT and MT evaluation. NIST open machine translation Evaluation series (OpenMT) was one of those events continuous from 2001to 2009 (Group, 2010). the annual Workshop on Statistical Machine Translation(WMT) held by the special interest group in machine translation (SIGMT) was continuing from 2006 to 2015 (Koehn and Monz, 2006; Callison-Burch et al., 2007; Bojar et al., 2015).

II. Methodology

Up to date and even for near future machine translation couldn't become an alternative for human translation for any translation domain such education, news, business, ..etc. but it has many merits over human translation such as cost, time saving and so on. Important question related to machine translation as Lavie et al.,(2010) state that 'For what purpose the MT output will be used' [3], anyway output quality should be considered for machine translation, in this view point we evaluated inkurdish machine translation system using Asiya toolkit.

As a part of the evaluation for the data human translation is used as a reference by the researchers and its meant to be preferred but the evaluation should be done by human and it's a shortcoming as (Zaretskaya, et al.,2016) declares that Human evaluation, on the other hand, is 'time-consuming and expensive, as well as subjective. We suggest that human evaluation procedure could be improved in order to achieve better efficiency and objectivity by developing a quantitative metric based on quality parameters and standards used in translation industry to evaluate human translation'. So, in order to avoid only human evaluation a well-known toolkit has been used as an evaluation toolkit with different metrics and is called Asiya toolkit.

As Ali Darwish (2001) cited in (Zaretskaya, et al.,2016) proposes a scale that also follows this schema. 'In his model for assessing translator's competence and the quality of translation as a product, he argues that each translation should be evaluated with regards to the purpose of translation, be it communicative, literal, reader-centered, and so on. He distinguishes the two attributes of translation: information integrity (which we call fidelity), linguistic integrity (which we call fluency)'. We can summarize our methodology in six steps:

1) Data set: we collected 50 different English samples for each of nouns, verbs, adjectives, uncommon words, daily expressions, sentences, idioms, proverbs, and paragraphs.
2) We got the reference translation (Kurdish equivalence) for all the data set by a Kurdish native – English Specialist Instructor.
3) We fed all English data set to 'inkurdish' machine translation system and got their Kurdish equivalence.
4) We fed Asiya toolkit with source text, machine output, and reference text to give us the evaluation result based on metrics we selected (BLEU,NIST, METEOR,-TER).
5) Calculating the average score for each data set using the formula :

   Average score = summations of all individual score / total number of data set.
6) Linguistic and computational analysis of the results.

A. Asiya Platform

For evaluating machine output there are many metrics such as BLEU, NIST,METEOR,..etc, majority of metrics with variety of options are grouped together under a project named Asiya which simplifies evaluation task, for this purpose we selected it .Asiya is an open Toolkit for machine Translation Evaluation which allows us to obtain automatic evaluation scores according to a selected set of metric representatives. Then, we can analyze our translations using the tools provided, such as the interactive plots and the automatic linguistic annotations(Meritxell Gonz_alez, et al., 2014).

A complete overview of the application architecture of Asiya and its modules is shown in Figure 1.

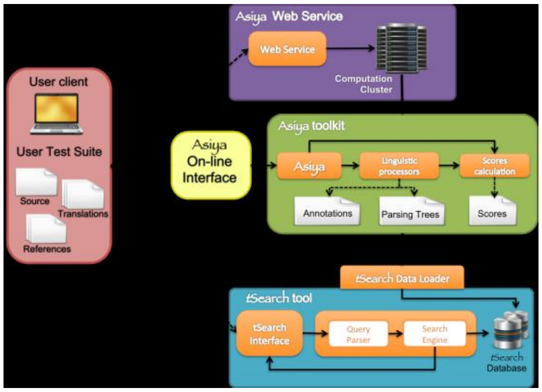


Figure 1: The Asiya Platform

B. Sample Testbeds

We started evaluating inkurdish from simple words including nouns, verbs, adjectives, and uncommon words in order to test the corpus behind the system,

Nouns	Lecturer :وانەبێژ
	Guidance : ڕێبەری
	Meat ball : شفتە
Verbs	Ruin : وێران کردن
	Slurp : لوشکە کردن
Adjectives	Sweet : شیرین
	Snobby : فشەکەر
Uncommon words	Lad : کوڕی گەنج
	Brisk : بروسکە

Table1: Sample of human translated words

Nouns	Lecturer :وانەبێژ
	Guidance : ڕێنمایی
	Meat ball : تۆپی گۆشت
Verbs	Ruin : وێران کردن
	Slurp : سلێپ
Adjectives	Sweet : شیرین
	Snobby : سنۆبی
Uncommon words	Lad : کوڕ
	Brisk : چالاک

Table2: Sample of machine translated words

From simple words we shifted to evaluate the system against simple sentence, compound sentence, proverbs, and idioms. Then, we have tested the system using daily expressions:

Daily expressions	how do you do چۆنی
	fair enough خراپ نیە
Sentences	I try to learn French. من هەوڵدەدەم فێری فەرەنسی ببم
	I had an accident. من توشی ڕووداوێک بووم
Proverbs	What glitters is not gold. هەرچی بدرەوشێتەوە ئاڵتوون نیە
	Like mother like daughter. دایک ببینە و کچ بخوازە
idioms	No pain, no gain. بە بێ ڕەنج ناگەیتە ئامانج
	Saving for a rainy day. ماڵی سپی بۆ ڕۆژی ڕەش

Table3: Sample of human translated Sentences

Daily expressions	how do you do تۆ چۆنی
	fair enough دادپەروەرانە بەس
Sentences	I try to learn French. من هەوڵدەدەم تا فێری فەرەنسی ببم
	I had an accident. من ڕووداوێکم تووش بوو
Proverbs	What glitters is not gold. چی درەوشانەوە زێڕ نیە
	Like mother like daughter. وەك دایك وەك كچ
idioms	No pain, no gain. هیچ ئازار، هیچ دەستکەوت
	Saving for a rainy day. ڕزگار کردن بۆ ڕۆژێکی باراناوی

Table4: Sample of machine translated Sentences

Finally, we have given the system some chosen paragraphs in different domains, here is an example about the movie:

I watched a horror film last night, it was very terrifying and I stop watching such movies. I always liked to watch horror films but I regret it now. I never recommend it to anyone else especially children.

Human Translation:

دوێنێ شەو سەیری فیلمێکی ترسناکم کرد، زۆر ترسناک بوو ئتر سەیری
ئەو جۆرە فیلمانە ناکەم .هەمیشە حەزم لە سەیر کردنی فلیمی ترسناک
بووە بەڵام ئێستا پەشیمانم.هەرگیز پێشنیاری ناکەم بۆ هیچ کەسێک بە
تایبەتی منداڵان .

Machine Translation:

من دوێنێ شەو فلیمێکی ترسم تەماشا کرد، ئەو زۆر دەتۆقاندی و من لە
تەماشا کردن فلیمی ئەوها دەوەستم. من هەمیشە حەزم کرد تا فلیمی ترس
تەماشا بکەم بەڵام من ئێستا ئەو پەشیمان دەبمەوە. من هەرگیز ئەو ڕا
ناسپێرم بۆ کەس هی تر بەتایبەتی منداڵ.

III. Result And Discussion

In this section we displayed the result of our work and discussed the result for each data fed to the inkurdish machine translation system as well as outputs getting from it using Asiya Toolkit for evaluation , finally we analyzed the result manually in addition to automatic evaluation done by Asiya toolkit in order to conclude the work in the best case.

A. Metric Set

As (Meritxell Gonz_alez,et al.,2014) stated that 'Asiya has a rich set of measures which evaluate translation quality based on different viewpoints and similarity assumptions. It has borrowed existing measures and has also implemented new ones' .

Asiya provide a description of the metric set. Metrics are grouped according to lexical, syntactic, and semantic level they operate. Common metrics measure the overlap in words and word sequences, as well as word order and edit distance (Aaron el at 2017).

Edit Distance: By calculating the minimum number of editing steps to transform output to reference (Aaron el at 2017), (Snover et al., 2006) design the translation edit rate (TER) (Translation Edit Rate) which measures the amount of post-editing necessary for machine output to be exactly matches a reference translation, TER has many variants but we selected the default one (-TER) which allows stemming and synonymy lookup but without paraphrase support. The TER score is calculated as:

$$\boxed{\mathrm{TER} = \frac{\#\text{of edit}}{\#\text{of average reference words}}} \quad (1)$$

Lexical Precision: (Papineni et al., 2002) designed the widely used evaluation metric BLEU. BLEU scores for several n-gram lengths (default = 4). The BLEU score is calculated as:

$$\boxed{\mathrm{BLEU} = \mathrm{BP} \times \exp \sum_{n=1}^{N} \lambda_n \log \mathrm{Precision}_n} \quad (2)$$

(Doddington, 2002) proposes the NIST metric as an enhancement of BLEU. NIST scores for several n-gram lengths (default = 5).

METEOR: (Banerjee and Lavie, 2005) design a novel evaluation metric METEOR. METEOR is based on general concept of flexible unigram matching, unigram precision and unigram recall. The METEOR score is calculated as:

$$\boxed{\begin{aligned} \mathrm{Penalty} &= 0.5 \times \left(\frac{\#\mathrm{chunks}}{\#\mathrm{matched\ unigrams}}\right)^3, \\ \mathrm{MEREOR} &= \frac{10PR}{R+9P} \times (1 - \mathrm{Penalty}). \end{aligned}} \quad (3)$$

B. Linguistic Analysis:

According to the sample test beds which are chosen among a huge number of data that can be found in the appendix, there are obviously advantages and disadvantages of inKurdish machine translation system such as the followings:

In this MT, there is another big problem, such as having no sound transcription to show the right pronunciation of the given words and instead it has the transcription in Kurdish script which is not accurate.

1. I try to learn French → من هەوڵ دەدەم تا فەڕەنسی فێر ببم

2. Cleanliness is next to godliness → پاکی لە تەنیشت خوناسیە

From the examples, we can understand the real problem of sentence structure which is a valid problem in inKurdish MT system. Because the words and their meanings are clarified in the corpus as shown in the above table of the MT but the synthesis and the structure of the combination of words are problematic. From the examples we can also understand that inKurdish MT system has different vocabularies in the corpus but it can't easily synthesize during translation due to lack of proper linking.

C. General Evaluation of inKurdish MT System

Lack of awareness of the culture is another big issue in translating idioms, proverbs and daily expressions with machine translation systems. Through different examples it is shown that there are several problems which inKurdish MT makes in translating English into Kurdish proverbs, idioms, paragraphs and daily expressions, such as(how is it going, is translated as چۆن ئەو دەڕوات which is translated literally which looks weird as (how is s/he going) and also the word (imposter) is only translated in a way which is not the only meaning of the word. The following sentence has been translated in the inkurdish in a very odd way with adding some words to the Kurdish translation which is not found in the English text (I like to learn English because it's an international language. I like to speak English everyday): من حەز دەکەم تا ئینگلیزی فێر ببم چونکەی زمانێکی نێودەوڵەتی. من حەز دەکەم تا ئینگلیزی ڕۆژانە قسە بکەم. So, adding something or clipping something else is very problematic in the English into Kurdish translation which can be seen in the inKurdish MT.

Moreover, all other examples from the Sample Table Test Beds clarify the fact that sentences can be understanding when translated in the machine but they are a bit odd in structure. Especially for Idioms and proverbs Human translation is more preferred compared to machine translation because there are different cultural references or understandings that can't be found in this MT and in MTs in general. So, proverbs, Idioms and daily expression can't get translated by these machine translators they way they are meant. This feature observed as a deficiency of Machine Translation Systems. However, there are other sentences included to this MT system and they have been translated in a good way as far as clear from the sample, when the sentence is very ordinary it can get translated easily but when its getting a bit complicated it can't get translated by this MT easily and quite meaningfully.

D. General Comments on Kurdish language

Kurdish owns some features which might be difficult to help an MT to work properly in terms of structure and the combination. Thus, these features of Kurdish are difficult to be investigated such as:

Firstly, having the feature of pro-drop property (means the subject is null and it can be dropped). Since the subject can be dropped in Kurdish and it can never be dropped in English might cause ambiguity in synthesizing a translated proper sentence from English into Kurdish or vice versa.

Secondly, having an SOV order where in every sentence their position is fixed and it has a very apparent different word order with English might cause difficulty to the system to switch the order between two different languages.

IV. FURTHER DISCUSSION

At a basic level for testing inkurdish machine translation system against single words (nouns, verbs, adjectives, and uncommon words) the result is predictable and acceptable except in few situations:

Example 1:

English Noun → Cashew

Reference → گازۆ

Inkurdish →داری کاشوو

Example 2:

English Noun →Meat Ball

Reference → شفتە

Inkurdish → تۆپی گۆشت

As we observed from both examples, the corpus which inkurdish relies on has lack of uncommon or less frequently used words as happened with the word ‘meat ball’, this leads to the fact that translation quality somehow depends on corpus quality.

At a sentence level, we have given the system many different sentences and adding words or clipping words happens in the translated sentences, two examples are just shown here:

Example 1:

Source text → I try to learn French.

Reference text → من هەوڵدەدەم فێری فەرەنسی ببم

Inkurdish text → من هەوڵ دەدەم تا فەڕەنسی فێر ببم

Example 2:

Source text → I had an accident.

Reference text → من توشی ڕووداوێک بووم

Inkurdish text → من ڕووداوێکم تووش بوو

The simple sentences we have translated with inkurdish system and the results we got from Asiya Toolkit tell us that the system relatively has no problem with translating simple English sentences into Kurdish as occurred with ‘I try to learn French’ which got near to 0.6 with BLEU metric .

At a proverb level, the system showed great shortcomings in its corpus, two examples are given:

Example 1:

Source proverb → what glitters is not gold.

Reference translation → هەرچی بدرەوشێتەوە ئاڵتوون نیە

Inkurdish translation →چی درەوشانەوە زێڕ نیە

Example 2:

Source proverb → like mother like daughter.

Reference translation → دایک ببینە و کچە بخوازە

Inkurdish translation → وەك دایك وەك کچ

One of the worst points we noticed with inkurdish system is dealing with proverbs, proverb handling is the simplest task in machine translation process, in simple way just saving the most commonly used English proverbs and giving their Kurdish equivalents. Inkurdish Proverb translation recorded very low scores near to 0.0875 with BLEU and 1- with PERbase metrics (Table 5).

Regarding idioms also two examples are given here,

Example 1:

Source idiom → no pain, no gain.

Reference equivalent → بە بێ رەنج ناگەیتە ئامانج

Inkurdish equivalent → هیچ ئازار، هیچ دەستکەوت

Example 2:

Source idiom → the ball is in your court.

Reference equivalent → بریار لای تۆیە

Inkurdish equivalent → تۆپەکە لە دادگاتە

Inkurdish dealing with idioms a little better than dealing with proverbs, idioms in inkurdish translation could be acceptable if and only if with post editing. Two examples regarding daily expressions are given here:

Example 1:

Source sentence → how do you do

Inkurdish translation → تۆ چۆنیت؟

Reference translation → چۆنی؟

Example 2:

Source sentence → so far so good

Inkurdish translation → تا ئێستا ئەوەندە باش

Reference translation → هەموو شتێک باشە

Inkurdish situation with daily expressions is better than proverbs and idioms but still has shortcoming that inkurdish couldn't involve at least the most commonly used daily expressions. Average score for inkurdish output for daily expressions is about 0.2 for BLEU metric (Table 5).

At a paragraph level, we have chosen different paragraphs from different domains such as science, literature, social, language and movie. Here we have shown just an example:

Source paragraph:

I watched a horror film last night, it was very terrifying and I stop watching such movies. I always liked to watch horror films but I regret it now. I never recommend it to anyone else especially children.

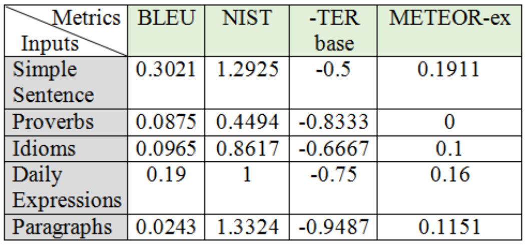
Reference translation:

دوێنێ شەو سەیری فیلمێکی ترسناکم کرد، زۆر ترسناک بوو ئتر سەیری ئەو جۆرە فیلمانە ناکەم .هەمیشە حەزم لە سەیر کردنی فلیمی ترسناک بووە بەڵام ئێستا پەشیمانم.هەرگیز پێشنیاری ناکەم بۆ هیچ کەسێک بە تایبەتی منداڵان .

Inkurdish output:

من دوێنێ شەو فلیمێکی ترسم تەماشا کرد، ئەو زۆر دەتۆقاندی و من لە تەماشا کردن فلیمی ئەوها دەوەستم. من هەمیشە حەزم کرد تا فلیمی ترس تەماشا بکەم بەڵام من ئێستا ئەو پەشیمان دەبمەوە. من هەرگیز ئەو ڕا ناسپێرم بۆ کەس هی تر بەتایبەتی منداڵ.

The problem of inkurdish in dealing with paragraphs is translating each sentence separately and isn't capable of linking a sentence to pre or next sentence, therefore pre editing in addition to post editing is required while translating a paragraph or a document using inkurdish machine translation system.



Metrics / Inputs	BLEU	NIST	-TER base	METEOR-ex
Simple Sentence	0.3021	1.2925	-0.5	0.1911
Proverbs	0.0875	0.4494	-0.8333	0
Idioms	0.0965	0.8617	-0.6667	0.1
Daily Expressions	0.19	1	-0.75	0.16
Paragraphs	0.0243	1.3324	-0.9487	0.1151

Table 5: Average Scores for different inputs

BLEU metric(score between 0 and 1) as declared by (Mohammed N. Al-Kabi) is based on counting the number of common words in the candidate translation and the reference translation, and then divides the number of common words by the total number of words in the candidate translation. NIST represents an enhancement to BLEU. TER(highest score is 0) tells us about the amount of post editing required for making the machine output closer to the reference translation while Meteor metric(score between 0 and 1) searches for exact matching. Table6 indicates that simple sentence scored highest among all data set (0.3021 out of 1) for BLEU metric while the lowest score is calculated for paragraphs, this shows the great shortcoming for 'inkurdish' system to deal with paragraphs and the reason is 'inkurdish' incapability to link among the sentences. Table6 also clarifies the amount of required post editing through –TER metric, both simple and compound sentences need the least amount of post editing (0.5) while most post editing amount required for both idioms and paragraphs(0.9487) . Exact matching between reference and machine translation calculated through Meteor metric, the worst score calculated for proverbs (0), zero means ‘inkurdish’ ignored proverb dealing in its corpus, this problem’s solution is quite easy by maintain the popular or most commonly used English proverbs as well as their Kurdish equivalences in the corpus. Meteor score for compound sentence is higher than simple sentence, this means that ‘inkurdish’ has better dealing with compound sentences compare to simple sentences; we clearly noticed this point while experimenting ‘inkurdish’ system and was unpredictable.

V. Conclusion

inkurdish as a starting point was an appreciable project but suffering from sharp shortcomings we concluded in this work, these are :

1) lack of rich corpus which involves all terms, synonyms, proverbs, idioms, and daily expressions.
2) Selecting suitable search algorithm that returns back an appropriate result is another demands should be considered for machine translation task but not found with inkurdish output.
3) Lack of using NLP techniques for getting proper instead of literal translation.
4) Facing problems with adding some words to the translated sentences or clipping some words from the source sentences, which causes inaccuracy in the translating process.

References

[Alon Lavie el at.,2010] Alon Lavie, ‘Evaluating the output of machine translation systems’, AMTA 2010 Tutorial.

[Banerjee and Lavie2005] Satanjeev Banerjee and AlonLavie. 2005. Meteor: An automatic metric formt evaluationwith improved correlationwith humanjudgments. In *Proceedings of the ACL*.

[Bojar et al.2015] Ondˇrej Bojar, Rajen Chatterjee, Christian Federmann, Barry Haddow, Matthias Huck, Chris Hokamp, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Carolina Scarton, Lucia Specia, and Marco Turchi. 2015. Findings of the 2015 workshop on statistical machine translation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 1–46, Lisbon, Portugal, September. Association for Computational Linguistics.

[Callison-Burch et al.2007] Chris Callison-Burch, Cameron Fordyce, Philipp Koehn, Christof Monz,and Josh Schroeder. 2007. (meta-) evaluation of machine translation. In *Proceedings of WMT*.

[Chomsky el at., 1988] Cook, V.J.' Chomsky's Universal Grammar'. In D. Crystal and K. Johnson(ed.) Applied Language Studies. Oxford: Basil Blackwell.1-196, 1988.

[Doddington2002] George Doddington. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *HLT Proceedings*.

[Faraj el at.,2009] Faraj, Hadar. ‘*Object Movement in Standard English and Central Kurdish’*. Unpublished MA, University of Koya, 2009.

[Group2010] NIST Multimodal Information Group, ‘Nist 2005 open machine translation (openmt) evaluation’. In *Philadelphia: Linguistic Data Consortium.Report, 2010*.

[Hashem el at., 2017] Hashm Ameen, unpublished MA. Thesis ‘Zmany farmy u pegay zmany kurdy la new zmana jihanyakanda’, Tahran university, 2017.

[Jan Berka el at.,2011] Jan Berka, Martin Černý, Ondřej Bojar, ‘Quiz-Based Evaluation of Machine Translation’, April 2011, 77–86.

[Koehn and Monz2006] Philipp Koehn and Christof Monz. 2006. Manual and automatic evaluation of machine translation between european languages. In *Proceedings on the Workshop on Statistical Machine Translation*, pages 102–121, New York City, June. Association for Computational Linguistics.

[Liu et al.2011] Chang Liu, Daniel Dahlmeier, and Hwee Tou Ng. ‘Better evaluation metrics lead to better machine translation’. In *Proceedings of EMNLP, 2011*.

[Mahwi el at.,2011] Mahwi, ‘*Bnamakany Sintaksy Kurdy’ (The principles of Kurdish Syntax)*. Slemany: Zankoy Slemany, M. 2011.

[Mohammed el at.,2013] Mohammed N. Al-Kabi, Taghreed M. Hailat, Emad M. Al-Shawakfa, and Izzat M. Alsmadi, ‘Evaluating English to Arabic Machine Translation Using BLEU’, Vol. 4, No.1, 2013, ppt. 66-73.

[Papineni et al.2002] Kishore Papineni, Salim Roukos,Todd Ward, and Wei Jing Zhu. 2002. Bleu: amethod for automatic evaluation of machine translation. In *Proceedings of ACL*.

[Qader el at.,2011] Qader, Tara. 2011. ‘*Zmany Kurdy u Minimal Program’ (Kurdish language and Minimalist Program)*, Unpublished PhD, University of Sulaimania.

[Snover et al.2006] Mattthew Snover, Bonnie J. Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. 2006. A study of translation edit rate with targeted human annotation. In *Proceeding of AMTA*.

[Stephen el at.,2010] Stephen Doherty, Sharon O’Brien, Michael Carl, ‘Eye tracking as an MT evaluation technique’, Springer Science+Business Media B.V. 2010.

[Wahab el at., 2015] Omer Fathulla Wahab, unpublished MA. ‘Thesis Pewazhokary khokaryanay zmany Kurdy’, Sulaimani university, 2015.