

English to Kurdish Rule-based Machine Translation System

Kanaan M. Kaka-Khan

Department of Computer Science, University of Human Development, Kurdistan Region, Iraq



ABSTRACT

Machine translation (MT) is a gaining ever more attention as a solution to overcome language barriers in information exchange and knowledge sharing. In this paper, we present a rule-based MT system developed to translate simple English sentences to Kurdish. The system is based on the Apertium free open-source engine that provides the environment and the required tools to develop a MT system. The developed system is used to translate some simple sentence, compound sentence, phrases, and idioms from English to Kurdish. The resulting translation is then evaluated manually for accuracy and completeness compared to the result produced by the popular (in Kurdish) English to Kurdish MT system. The result shows that our system is more accurate than in Kurdish system. This paper contributes toward the ongoing effort to achieve full machine-based translation in general and English to Kurdish MT in specific.

Index Terms: Apertium, Inkurdish, Machine Translation, Morphological, Rule-based Machine Translation

1. INTRODUCTION

This paper presents a rule-based machine translation (RBMT) system for the Kurdish language. The goals of this paper are two-fold: First, we build MT system using a free/open-source platform (Apertium). Second, we evaluate the translation of proposed system with “inkurdish” translation for the same set of data through manual evaluation method.

The Kurdish language belongs to the group of Indo-European languages. The Kurdish dialects are divided, according to the linguistic and geographical facts, into four main dialects. They are the North Kurmanji, Middle Kurmanji, South Kurmanji, and Gurani [1]. Kurdish is written using four different scripts, which are modified

Persian/Arabic, Latin, Yekgirtu (unified), and Cyrillic. Latin script uses a single character while Persian/Arabic and Yekgirtu in a few cases use two characters for one letter. The Persian/Arabic script is even more complex with its RTL and concatenated writing style [2].

MT, perhaps the earliest NLP application, is the translation of text units from one natural language to another using computers [3]. Achieving error-free translation is a difficult task, instead an improvement in completely automatic, high quality, and general-purpose translations is required. The better MT evaluation metrics will be surely helpful to the development of better MT systems [4]. The MT evaluation has both automatic and manual (human) evaluation methods; the human evaluation criteria include the fluency, adequacy, intelligibility, fidelity, informativeness, task-oriented measures, and post-editing. The automatic evaluation method criteria include precision, recall, F-measure, edit distance, word order, part of speech tag, sentence structures, phrase types, named entity, synonyms, paraphrase, semantic roles, and language models. For this work, manual evaluation method has been used to evaluate the accuracy of both the systems.

Access this article online

DOI: 10.21928/uhdjst.v2n2y2018.pp32-39	E-ISSN: 2521-4217 P-ISSN: 2521-4209
----------------------------------------	----------------------------------------

Copyright © 2018 Kaka-Khan. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Kanaan M. Kaka-Khan, Department of Computer Science, University of Human Development, Kurdistan Region, Iraq. E-mail kanaan.mikael@uhd.edu.iq

Received: 11-08-2018

Accepted: 17-08-2018

Published: 02-09-2018

We have used a platform called Apertium; Apertium defines itself as a free/open-source MT platform, initially aimed at related-language pairs but expanded to deal with more divergent language pairs and provide a language-independent MT engine and tools to manage the linguistic data [5].

Apertium originated as one of the MT engines in the project OpenTrad, which was funded by the Spanish government and developed by the Transducens research group at the Universitat d'Alacant. At present, Apertium has released 40 stable language pairs. Being an open-source project, Apertium provides tools for potential developers to build their own language pair and contribute to the project. Although Translators without Borders (TWB) claimed that they have developed offline MT engines for Sorani and Kurmanji, specifically for translating content for refugees using apertium, their work had not been published academically. Although Apertium was founded initially to provide an English/Catalan converter, it can also be used to right to left languages with more efforts specifically in creating transfer rules.

The rest of this paper is organized in the following way: Next, we present MT survey in Section 2. We describe methodology in Section 3. We then show and explain the results in Section 4, followed by the conclusion in the last section.

2. MT SURVEY

2.1. General MT Survey

A very early MT system returned to 1950s [6]. The development of computer with high storage and performance in one side and availability of bilingual and multilingual corpora in other side led to gain rapid MT development since the 1990s [7]. In 1993, IBM Watson research group did many important achievements in MT areas such as designing five statistical MT models and the techniques to estimate the model parameters using bilingual corpora [8]. In 2003, Franz Josef presented minimum error rate training for statistical MT systems [9] and Koehn *et al.* proposed statistical RBMT model [10]; in 2005, Koehn and Monz presented a shared task of building statistical MT systems for four European languages [11], and David Chiang proposed a hierarchical phrase-based SMT model that is learned from a bitext without syntactic information [12]; Menezes *et al.* used global reordering and dependency tree to build English-to-Spanish statistical MT in 2006 [13]. In 2007, Koehn *et al.* did a great achievement which was developing Moses, an open-source SMT software

toolkit [14]; at the same time, in the sake of improving word alignment and language model quality among different languages, Hwang *et al.* team utilized the shallow linguistic knowledge [15]; Sa´nchez-Mart´inez and Forcada described an unsupervised method for the automatic inference of structural transfer rules for a shallow-transfer MT system in 2009 [16]. In 2011, Khalilov and Fonollosa designed a new syntax-based reordering technique to determine the problem of word ordering [17].

Deep learning fast development played a great roles in MT research evolving from conventional models to example-based models by Nirenburg in 1989 [18], statistical models by Carl and Way in 2003 [19], hybrid models by Koehn and Knight in 2009 [20], and recent years' Neural models by Bahdanau *et al.* in 2014 [21].

Neural MT (NMT) is a recently hot topic that leads the automatic translation to be worked in a very different direction with the traditional phrase-based SMT methods. In traditional model, the different MT components are training separately, while the NMT components are training jointly by utilizing artificial neural network to increase the translation performance through two step recurrent neural network of encoder and decoder [21]-[23].

2.2. Kurdish MT Survey

Unfortunately, few efforts have been done for Kurdish MT yet. In 2011, Safeen Ghafour proposed a project called Speekulate; Speekulate can be considered as a theoretical research, a multiuse translator [24]. In 2013, the first English to Kurdish (sorani) MT system has been released under the name "inkurdish" for translating English text to Kurdish language [25]. In 2016, Google translate has added support for 13 new languages including Kurdish (Kurmanji dialect) language, bringing the total number of supported tongues to 10 [26]. TWB has developed offline MT engines for Sorani and Kurmanji, specifically for translating content for refugees [27]; in 2017, Kanaan and Fatima have evaluated "inkurdish" MT system using different automatic evaluation metrics in the sake of touching the weaknesses of "inkurdish" MT system [28]; Hassani suggested a method for MT among two Kurdish dialects (Kurmanji and Sorani) using bidialectal dictionaries, and his result showed that the translated texts are in 71% and 79% of cases rated as understandable for Kurmanji and Sorani, respectively. They are rated as slightly understandable in 29% of cases for Kurmanji and 21% for Sorani [2].

3. METHODOLOGY

The nature of language and availability of resources play important roles in selecting MT approach. Fig. 1 describes the four different categories of machine translation approaches.

3.1. Direct Translation

Direct translation involves a word-by-word translation approach. No intermediate representation is produced.

3.2. Rule-based Translation

RBMT systems parse the source text and produce an intermediate representation. The target language text is generated from the intermediate representation.

3.3. Corpus-based Translation

The advantages of this approach are that they are fully automatic and require less human labor. However, they require sentence-aligned parallel text for each language pair and cannot be used for language pairs, for which such corpora do not exist.

3.4. Knowledge-based Translation

This kind of system is concerted around “Concept” lexicon representation a domain.

Rule-based approach has been chosen for this proposed system; reasons to choose a rule-based instead of a statistic

system depend on the unavailability of sufficiently large corpora [29]; we use a RBMT which is suitable for languages, for which there are very little data [27]; despite being spoken by about 30 million people in different countries, Kurdish is among less-resourced languages [2]. Hence, RBMT is a suitable choice for Kurdish MT. RBMT models transform the input structure to produce a representation which matches the target language rules, and it has three components (Fig. 2): Analysis, to produce the structure of source language; transfer, to transfer the representation of source language to representation of a target language; and generation, using target level structure to generate target language text.

After completing the prototype of the system, 500 different random data sets (simple sentence, complex sentence, proverbs, idioms, and phrases) have been given to both systems. Then, the output of both systems has been given to an annotator (English specialist - Kurdish native), to evaluate the results through manual evaluation method. The aim of the evaluation is to determine the translation accuracy for both systems in both meaning and grammar correctness. The evaluation has been designed into 5 categories, from score 5–1: Highly accurate, the translation is very near to the reference, it conveys the content of the input sentence, and no post editing is required; accurate, the translation conveys the content of the input sentence, and little post-editing

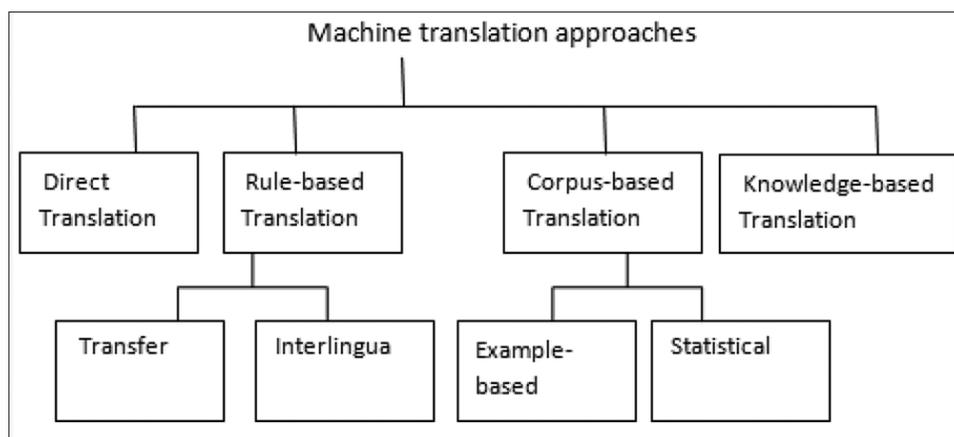


Fig. 1. Machine translation approaches [1].

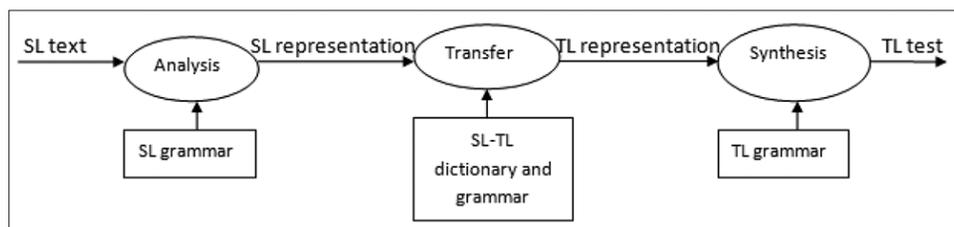


Fig. 2. Rule-based (transfer-based) machine translation diagram [2].

is required; fairly accurate, while the translation generally conveys the meaning of the input sentence, it suffers from word order problems or tense or un-translated words; poorly accurate, while the translation somehow conveys the meaning of the input sentence, it does not convey the input sentence content accurately; and completely inaccurate, the content of the input sentence is not conveyed at all by the translation, and it just give the translation of the words individually.

4. PROPOSED SYSTEM CONFIGURATION

Our system basically works on dictionaries and transfer rules, and at a basic level, we maintain three main dictionaries:

1. Kurdish morphological dictionary: This file describes the rules of how words in Kurdish language are inflected, and its named: Apertium-kur.kur.dix
2. English morphological dictionary: This file describes the rules of how words in English language are inflected, and its named: Apertium-eng.eng.dix
3. Bilingual Dictionary: This file describes correspondences between words and symbols in Kurdish and English languages, and its named: Apertium-kur-eng.kur-eng.dix.

We maintain files for transfer rules in the two languages. The rules govern the words reordering in target language, the file is:

- English to Kurdish language transfer rules: This file contains rules govern how English will be changed into Kurdish language, its named: Apertium-eng-kur.kur-eng.t1x.

In spite of the possibility of translating Kurdish to English texts, we just present English to Kurdish translation in this work.

4.1. Terms Used in the System

Before creating the dictionaries and rules, some related terms would be explained briefly. The first term is lemma: Lemma is the form of word which is stripped of any grammatical information, for example book is the lemma of (booked, booking, etc.) and be is the lemma of was. The second term is symbol: A grammatical label for example singular and plural names, first person and present indicative, etc. Tags are used for symbols, <n> for noun, <pl> for plural, etc. paradigm is the another related term which refers to inflection of a particular group of words, for example happy, happ (y, ier, iest), instead of storing a lot of the same thing, We can simply store one, and then we say the second inflects like the first, for example “shy, inflects like happy”. Paradigms are defined in <pardef> tags, and used in <par> tags.

4.2. Basic Tags in Kurdish and English Dictionaries

<Dictionary><dictionary/> tag is the start and end point which contains the other all tags within xml file. <Alphabet><alphabet/> tag defines the set of letters that will be used in the dictionary.

<alphabet>ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz<alphabet/> for English dictionary.

<alphabet>ئ- ا ب پ ت ج چ ح خ د ر ز و ه ن م ل گ ک ق ف غ ع ش س ژ ز ر ئ<alphabet/> for Kurdish dictionary.

Symbol definitions: The symbols name can be written out in full or in abbreviate, for example, noun (n) in singular (sg) and plural (pl) (Fig. 3).

Then, we define a section <section><section/> for the paradigms <pardefs><pardefs/> (Fig. 4).

This is the basic skeleton for the morphological dictionaries, then the words will be entered through the entries, <e><p><l/><r><s n="n"/><s n="sg"/></r></

```
<sdefs>
  <sdef n="/n"/>
  <sdef sg="sg"/>
  <sdef pl="pl"/>
<sdefs/>
```

Fig. 3. Tags used for symbols.

```
<? xml version="1.0" encoding="UTF-8?>
<dictionary>
  <sdefs>
    <sdef n="/n"/>
    <sdef sg="sg"/>
    <sdef pl="pl"/>
  <sdefs/>
  <pardefs>
  <pardefs/>.
  <section id="main" type="standard">
    <section/>
  </dictionary/>
```

Fig. 4. Skeleton for morphological dictionary.

p></e>, here e for entry, p for pair, l for left, and r for right. Compiling entries left to right lead to produce analyses from words and from right to left leads to produces words from analyses. The final step is compiling and run the dictionary. Both English (apertium-eng.eng.dix) and Kurdish (apertium-kur.kur.dix) morphological dictionaries would be created in the same manner.

4.3. Bilingual Dictionary

This describes mappings between words, the basic skeleton is the same as monolingual dictionary, but we need to add an entry to translate between the two words:

<e><p><l>university<s n="n"/><l/><r> وکەناز <s n="n"/></r></p></e>. We compile the bilingual dictionary left to right to produce the Kurdish→ English dictionary and right to left to produce the English → Kurdish dictionary.

4.4. Transfer Rules

It contains rules to govern how English will be changed into Kurdish language, and the basic skeleton of the transfer rules is shown here (Fig. 5).

<rule> tag defines a rule. <pattern> tag means: “Apply this rule, if this pattern is found” (Here the pattern consists of a single noun defined by the category item nom). Patterns are matched in a longest-match first. The pattern matched and rule executed would be the first one. For each pattern, there is an associated action, which produces an associated output, out. The output is a lexical unit (lu).The <clip> tag allows a user to select and manipulate attributes and parts of the source language (side="sl") or target language (side="tl") lexical item. Transfer rules file need to be compiled and tested.

5. RESULTS AND DISCUSSIONS

After completing the prototype of the proposed system, it would be tested against different sets of data; first, we test it against individual words, and then simple sentence, complex sentences, phrases, proverbs, and idioms, some examples are shown in Fig. 6.

Fig. 6 shows a random sample of data translated by our proposed system; we tried to maintain a rich corpus that involves vast numbers of individual words, phrases, idioms, proverbs, etc., in order not to have un translated words in the output.

The second part of this work will be evaluation between the proposed system’s results with “inkurdish” MT system

```
<rule>
  <pattern>
    <pattern-item n="nom"/>
  </pattern>
  <action>
    <out>
      <lu>
        <clip pos="1" side="tl" part="lem"/>
        <clip pos="1" side="tl" part="a_nom"/>
        <clip pos="1" side="tl" part="nbr"/>
      </lu>
    </out>
  </action>
</rule>
```

Fig. 5. Skeleton for transfer rules.

```
$ echo "University" | lt-proc kur-eng.automorf.bin
Output: ^ زانکو
$ echo "bird" | lt-proc kur-eng.automorf.bin
Output: ^ باندە
$ echo "white color" | lt-proc kur-eng.automorf.bin
Output: ^ ڕەنگی سپی
$ echo "He ate an apple" | lt-proc kur-eng.automorf.bin
Output: ^ ئەو سیوێکی خوارد
$ echo "he went to play football" | lt-proc kur-eng.automorf.bin
Output: ^ ئەو ڕۆشت یاری تۆپی پێ بکات
$ echo "she bought a new dress" | lt-proc kur-eng.automorf.bin
Output: ^ ئەو کراسێکی تازەى کرى
$ echo "while she was watching the movie, the power switched off" | lt-proc kur-eng.automorf.bin
Output: ^ کاتیگ تەمامایى فیلەمە کەمى دەکرد، کارەبا کوژایەوه
$ echo "saving for a rainy day" | lt-proc kur-eng.automorf.bin
Output: ^ مالى سەجى بۆ ڕۆژى ڕەش
$ echo "what glitters is not gold" | lt-proc kur-eng.automorf.bin
Output: ^ ئەوێى بێرەوشیپتەوه نائەون نیه
```

Fig. 6. Samples of proposed system translation.

results for the same set of data using manual evaluation method. Table_1 shows a sample of data translated by both systems. Inkurdish non-sense output with paragraphs and long texts obliged us to be satisfied at basic level (simple and compound sentence, idioms, proverbs, and phrases) evaluation; the sample contains a couple of random examples of each data set. The evaluation made by a neutral annotator (Kurdish native which is English specialist) according to the five categories has been defined before.

Detailed explanation of both computational and linguistics issues is out of our main aim, and we focused on accuracy differences between both systems, plus touching some general translation issues found during experimenting the data sets. Inkurdish MT system suffers severely from some issues, it is unable to link verbs to objects in sentences, and in spite of having all different meaning for a specific verb

TABLE 1
Sample of data sets with their translations

Dataset	Source text	Inkurdish translation	Proposed system translation
Simple sentence	I go to university	ۆكناز ۆب مۆرهد نم	ۆكناز ۆب مۆرهد نم
Evaluation		5	4
Evaluation	The kids are playing in the backyard	مكهدايكباب هل نهكهد یرای ناكهلادنم	یهچخاب هل نهكهد یرای ناكهلادنم هوششپ
Complex sentence	I have been playing football since I am 6 years old	6 جنم یهتهوهل ینیبهد لۆر مئپ یپۆت نم نالاس	یهتهوهل مکههئ یپۆت یرای نم نالاس 6 منمهت
Evaluation		3	4
Evaluation	While he was watching the movie, the power switched off	دركهد اشامت یهكهمیلف وهئ یهتاك وهل جوهیدن اژوك ان اوت	ایشامهت (مكهروك) وهئ یهتاك وهل ابهراك یوزمهت، دركهد یكهمیلف هوهی اژوك
Evaluation proverbs	Better late than never	زیگرهه هل گنمرد رتشیاب	نیشهگهن هل نیشهگ گنمرد رتشیاب
Evaluation		2	5
Evaluation Idioms	Actions speak louder than words	هشو هل نهكهد مسق رتزرهب یگنهد راک	مسق کهن متهرش رادرک
Evaluation		2	4
Evaluation	The English test was a piece of cake	یکهئیهچرپ یزیلگن یهئ یهمن درکیقات ووب کئیک	رۆز یزیلگن یهئ یهمن درک یقات ووب ناسائ
Evaluation		2	4
Evaluation phrases	You can kill two birds with one stone	لگهل تیزو کب مدن لآب وود تین اوتهد وت درهب کههئ	وود درهب کههئ هب تین اوتهد وت تیزو کب مدن لآب
Evaluation		2	3
Evaluation	Thanks I am pretty good	مشاب کههئ هدار ات نم سناپوس	مشاب رۆز نم سناپوس
Evaluation		3	5
Evaluation	We could have dinner at MacDonald. How does that sound?	دل نهؤدکهم هل نامین اوت وئش نیؤخب مهنئ؟ تاکهد هگنهد وهئ نۆچ	دل انؤدکام هل وئش نین اوتهد مهنئ؟ هگنهد وهئ هنۆچ، نیؤخب
Evaluation		2	3

in the corpus, it fails to bring the correct meaning of the verb according to its position in the sentence; it translated the verb “play” in “He went to play football before 1 h” example (Table 1) as “تاکهد یرای” instead of “تین یبهد لۆر” and this led to improper translation; the corpus of inkurdish suffers from lack of pre defined common English idioms and proverbs; it always gives literal translation for idioms and proverbs for example, it translated “Better late than never” proverb to “زیگرهه هل گنمرد رتشیاب” (Table 1) which is very literal and non-sense translation. Untranslated word is another issue for inkurdish system for example the word “backyard” has not been translated in “The kids are playing in the backyard” example (Table 1).

Table 2 shows the accuracy average for all different data sets of both systems, and the accuracy averages have been calculated through a simple formula: Average = summation of all individual scores/total number of samples. The results showed that our system is more accurate than inkurdish system for all data sets; both systems got high scores with “simple sentence” translation (3.12 and 3.56 of 5 for inkurdish and our system, respectively); inkurdish got the least score for idioms while our system for phrases (1.15 and

TABLE 2
Translation accuracy average for both systems

Dataset	Inkurdish accuracy average	Proposed system accuracy average
Simple sentence	3.12	3.56
Complex sentence	2.45	2.78
proverbs	1.25	2.22
Idioms	1.15	2.42
Phrases	1.46	2.13

2.13 of 5, respectively), this means that inkurdish needs to maintain large number of common English proverbs and idioms with their Kurdish equivalents while our system need to involve more English phrases.

In our previous work “Evaluation of inkurdish MT System,” we addressed the issues of this MT system in details; hence, we tried to bridge the gaps of inkurdish system in our proposed system and this is the reason of clear differences between inkurdish accuracy average and proposed system accuracy average; the most common inkurdish issue is lack of rich corpus specifically to deal with phrases, idioms, and proverbs (1.46, 1.15, and 1.25, respectively) (Table 2); during

experimenting the data with inkurdish, it did not translate even one idiom or proverb, and it gives a literal translation instead.

6. CONCLUSION

MT remains to be one of the most challenging aspects of NLP. Despite the ongoing efforts to achieve full machine-based translation, little progress has been achieved; due to language structure and composition complexity. Open-source platforms have provided the environment and tools required to develop reliable MT systems, especially for language with poor resources such as Kurdish. We have presented a MT system to translate English to Kurdish developed using an open-source platform. The resulting translation is compared with the result generated by inkurdish popular English to Kurdish MT system. The result shows clear differences between inkurdish MT system and our MT system in terms of translation accuracy. The result also shows that RBMT and manual MT evaluation are suitable choices, for poorly resourced languages.

Biography

Kanaan M.Kaka-Khan is an associate professor in the Computer Science department at Human Development University, Sulaimaniya, Iraq. Born in Iraq 1982. Kanaan M.Khan had his bachelor degree in Computer Science from Sulaimaniya University and Master Degree in IT from BAM University, India. His research interest area include: Natural Language Processing, MT, Chatbot, and Information Security.

REFERENCES

- [1] F. H. Khorshid. Kurdish Language and the Geographical Distribution of its Dialects. Ishbeelia Press, Baghdad, 1983.
- [2] H. Hassani. Kurdish Interdialect Machine Translation. in *Proc of the fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, 2017, pp. 63-72.
- [3] T. Siddiqui and U. S Tiwary. Natural Language Processing and Information Retrieval. Oxford University Press. Second Impression 2010, Oxford, 2010.
- [4] C. Liu, D. Dahlmeier and H. T. Ng. Better Evaluation Metrics Lead to Better Machine Translation, in: *Proc EMNLP*, 2011.
- [5] Apertuim. About Apertuim, 2018. Available: www.apertium.org/index.eng.html?dir=eng-spa#translation. [Aug. 7, 2018].
- [6] W. Weaver. Translation. Machine Translation of Languages: Fourteen Essays. MIT Press, Cambridge, MA, 1955.
- [7] J. B. Marin^o, R. E. Banchs, J. M. Crego, A. Gispert, P. Lambert, J. A. R. Fonollosa and M. R. Costa-jussa¹. N-gram based machine translation. *Computational Linguistics*, vol. 32, no. 4, pp. 527-549, 2006.
- [8] P. F. Brown, V. J. D. Pietra, S. A. D. Pietra and R. L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, vol. 19, no. 2, pp. 263-311, 1993.
- [9] F. J. Och. Minimum error rate training for statistical machine translation. in *Proc. of ACL*, 2003.
- [10] P. Koehn, F. J. Och and D. Marcu. Statistical phrase-based translation. in *Proc. of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Association for Computational Linguistics*. vol. 1, pp. 48-54, 2003.
- [11] P. Koehn and C. Monz. Shared task: Statistical machine translation between European languages. in *Proc. of the ACL Workshop on Building and Using Parallel Texts*, 2005.
- [12] D. Chiang. A hierarchical phrase-based model for statistical machine translation. in *Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2005, pp. 263-270.
- [13] A. Menezes, K. Toutanova and C. Quirk. Microsoft research treelet translation system: NaacL 2006 europarl evaluation. in *Proc. of WMT*, 2006.
- [14] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. J. Dyer, O. Bojar, A. Constantin and E. Herbst. Moses: Open source toolkit for statistical machine translation. in *Proc. of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, Association for Computational Linguistics*, 2007b, pp. 177-180.
- [15] Y. S. Hwang, A. Finch and Y. Sasaki. Improving statistical machine translation using shallow linguistic knowledge. *Computer Speech and Language*, vol. 21, no. 2, pp. 350-372.
- [16] F. SaⁿchezMartⁱnez and M. L. Forcada. Inferring shallow-transfer machine translation rules from small parallel corpora. *Journal of Artificial Intelligence Research*, vol. 34, pp. 605-635, 2009.
- [17] M. Khalilov and J. A. R. Fonollosa. Syntax-based reordering for statistical machine translation. *Computer Speech and Language*, vol. 25, no. 4, 761-788, 2011.
- [18] S. Nirenburg. Knowledge based machine translation. *Machine Translation*, vol. 4, no. 1, pp. 5-24, 1989.
- [19] M. Carl and A. Way. Recent Advances in Example-Based Machine Translation. Kluwer Acadmic Publishers, Dordrecht/Boston/London, 2003.
- [20] P. Koehn and K. Knight. Statistical Machine Translation, November 24. US Patent No. 7,624,005, 2009.
- [21] D. Bahdanau, K. Cho and Y. Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, vol. abs/1409.0473, p. 9, 2014.
- [22] K. H. Cho, B. van Merriënboer, D. Bahdanau and Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, vol. abs/1409.1259, p.15, 2014.
- [23] K. Wolk and K. Marasek. Neural-based machine translation for medical text domain. Based on European medicines agency leaflet texts. *Procedia Computer Science*, vol. 64, p. 2-9. 2015.
- [24] S. Ghafour. "Speeculate". 2011. Available: www.kurditgroup.org/sites/default/files/speeculate_0.pdf. [Aug. 4, 2018].
- [25] Inkurdish Translator. 2013. Available: www.inkurdish.com/. [Aug. 4, 2018].
- [26] Ekurd Daily-Editorial Staff. "Google Translate adds support for Kurdish language", 2016. Available: www.ekurd.net/google-translate-kurdish-language-2016-02-18. [Aug. 4, 2018].
- [27] Translators Without Borders. "Translators without Borders

developed the world's first crisis-specific machine translation system for Kurdish languages", 2016. Available: www.translatorswithoutborders.org/translators-without-borders-develops-worlds-first-crisis-specific-machine-translation-system-kurdish-languages/. [Aug. 4, 2018].

[28] K. M. Kaka-kHan and F. Jalal. Evaluation of In Kurdish machine

translation system. Presented in the 4th international scientific conference of university of human development, Apr. 2017. *Journal of University of Human Development*, vol. 3, no. 2, pp. 862-868, Jun. 2017.

[29] W. Linda. Rule-based MT Approaches Such as Apertium and GramTrans. Universitetet i Tromsø, Norway 21.10.2008, 2008.