

Linguistic Obstacles in Machine Translation: English and Kurdish Language

Khoshy Anwar Hassan^{1*} Shokhan Anwar Hassan²

¹Department of English Language, College of Education- University of Garmian-Kalar. Kurdistan Region, Iraq.

²Kalar Computer Institute. Ministry of Education-General Directorate of Institutes and Trainings. Kurdistan Region, Iraq.

E-mail: shokhan.anwar@gmail.com

* Corresponding author. E-mail: khoshy.anwar@garmian.edu.krd

Abstract

Machine Translation (MT) is an application of Natural Language Processing (NLP) which is an application of Artificial Intelligence (AI). NLP combines both Computer Science and Linguistics fields. One of the reasons for MT that should be taken into consideration is challenges for traditional translations including cost, time wasting...etc. Although Machine Translation nowadays has important roles in translation but still it cannot be considered as an alternative for traditional translation completely. Unfortunately, as Kurdish nation, we lack having a machine translation system capable to translate Kurdish sentences perfectly. Thus, this paper tackles the major problems or obstacles for Kurdish machine translation in detail. It can be regarded as a basic or fundamental point for anyone who would like to work in this area.

Keywords: Machine Translation, Kurdish language, Phrase translation, Structures, Parts of speech, Linguistic challenges.

1. Introduction

In dealing with the computational Linguistics, the first thing that people should do is to try all their bests in understanding the nature of the language and the rules, which human language operates. This is to find the mechanism of this operation and then to simulate it through the automatic means. While translating, human beings usually try to clarify the source language on three levels:

Semantic level: understanding words out of context, as in a dictionary.

Syntactic level: understanding words in a sentence.

Pragmatic level: understanding words in situations and context.

In order to have a high quality output, the same idea should be processed while doing automatic translation (machine translation).

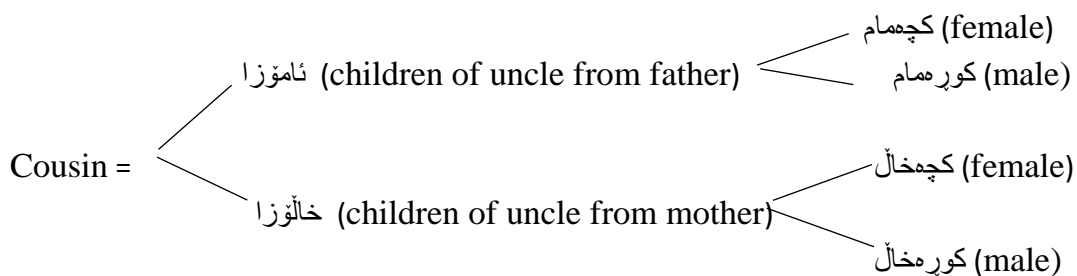
2. Obstacles and Issues

Word Equivalency between the two languages

In some cases, a word in Kurdish language may not have an equivalent word meaning in another language such as English or Arabic...etc. Instead it may be expressed by a group of words in English or vice versa. Kurdish-English example:

Kurdish	English
كونه	Water storage that made of animal skin, specially lamb and sheep
مهشكه	Sour milk storage that made of animal skin, specially lamb and sheep

Salam (no year: 96) identifies that Kurdish language has more terms regarding family relations than English and Arabic. For example, in English language the word (cousin) is a generalized word while in Kurdish language it is specified also gender considered. As in:



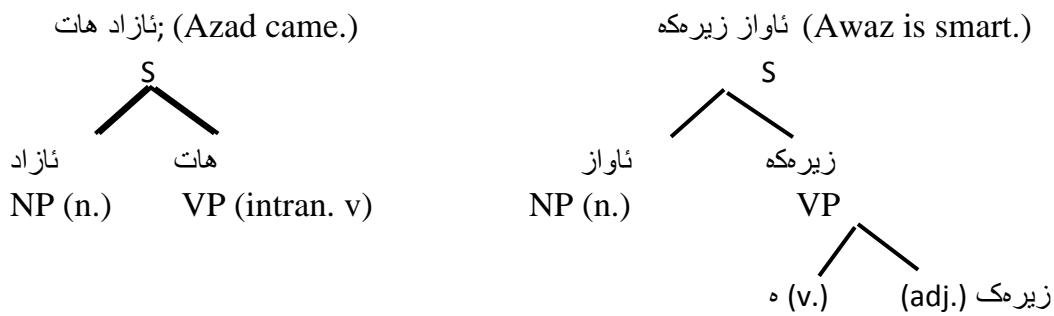
Phrase Translation

One of the problems, in traditional or direct oral translation, is that "not all phrases in one language have equivalent phrases in another language" in such cases, dictionary translation would not provide a proper or meaningful translation. For instance; in the case of verb phrase (VP) as Crystal (2003:352) defines verb phrase as "a type of verb consisting of a sequence of a lexical element plus one or more particles", Kurdish VP contains only one main verb while in English, it is two parts which are auxiliary/modal and main verb. In addition, Kurdish linguists agreed on defining a sentence as "a sequence of words that provide a full perfect meaning that has a symbol at the end" (Wirya, 2011 A: 34). While Nariman (2012: 13-14) rejects that definition to be generalized in all sentences of Kurdish language since there should be a sentence consisting of a phrase alone as in:

مرد. = (s/he passed away)

S→VP→v.

He also says that the rule of "S= NP + VP" cannot be applied in all Kurdish sentences, while this tackles only those with non-transitive verbs or those with "weak" verb. As in:



Zare (2005: 98) explains another problematic issue for machine translation. She states that Noun phrase in Kurdish language can be modified or post modified by ordinal number and still with the same meaning, while in English language it is not acceptable to be post modified by ordinal it is rather to be cardinal number.

وانه‌ی یه‌که‌م = Lesson one

یه‌که‌م وانه = First lesson

Nariman (2010: 131), and Zare (2005: 184) provide that there is a problem with Kurdish noun phrase as compared to English. The cardinal numbers that pre-modifies the noun will not pluralize the noun, For example; یه‌ک = one, سه‌ی = three.

(یه‌ک پیاو) هات.

(سه‌ی پیاو) هاتن.

سه‌ی که‌س چوونه شار. Three persons went to the city.

While in English the noun will be pluralized as in:

He has (one sister) and (three brothers) (Quirk and Greenbaum, 1973: 65).

Different Structures

English has SVO structure while Kurdish sentence structure is differ. Wirya (2011 A: 35-36) and Omed (2011: 135-7) states that Kurdish has SOV structure. Yet, in some cases OSV, OVS, VSO, and SVO can also be found in the surface structure that are derived from the same deep structure by the application of different transformational rules. As shown in the following sentences:

(SOV) نازاد ئهم گوله‌ی کړی.

(OSV) ئهم گوله‌ی نازاد کړی.

(OVS) ئهم گوله‌ی کړی نازاد.

In the verb phrase where the subject and object are represented by personal suffixes the orders of VSO, SVO and OVS are found.

بینی-م-یت. (VSO) = I saw you

ئه-تان-ناردن. (SVO) = you used to send them

ئه-یان-نیرم. (OVS) = I send them.

Kurdish is an agglutinative language while English is not since its patterns are separate (Wirya, 2011 B: 123)

Moreover, Nariman (2012: 14) says that the definition (Kurdish sentence contains subject, direct object and a verb) is not applicable for all Kurdish sentences. It only tackles those that their verbs are transitive, as in:

کارزان وانهکهی خویند. (SOV) = Karzan studied the subject.

Rasul (2005: 192) highlights that in Kurdish language there is transformational rule that makes the syntactic items change their location. Although Kurdish is an SOV language, but because of those movements, sometimes the verb can come in the beginning as in:

چوو-م بۆ بازار. (VSO)

These changes break the grammatical rules, and usually the poets play with the items for the sake of their poem. Omer (2015:56) in his research explains that there are mismatches in languages, which is very hard for machine translation to perform its work. As in the case of idioms, not only the translation should be using different words while their cultural symbols and meanings also do not match. For instance:

تهور و داس دهباریت. (It is raining axe and billhook.)

in English the expression is different which means (It is raining cats and dogs)

پشیلهکان و سهگهکان دهبارین

So, in translation, the above structures of the Kurdish language should be well defined. As for passive form the elements of Kurdish language do not match with English for translation. Sajida (2013: 38) clarifies the process of passive form in Kurdish language as (SOV=1, 2, 3) in the active form, when it changes to passive it will be (OV=2, 3). This process changes in English sentence, as Leech et al. (2001: 363) simplify passive form in English language as (SVO=1, 2, 3) is the active form, transferring it to passive the result will be (OV=3, 2). The following examples provide more detail about changing sentences from active to passive form:

من خانوو مکهم فرۆشت.

(3) (2) (1)

خانوو مکه فرۆشرا.

The cat chased the mouse.

(1) (2) (3)

The mouse was chased.

Parts Of Speech Equivalent

English POS may not have correspondence Kurdish POS. For Omer (2015: 57-58), the issue in Kurdish machine translation is that parts of speech in Kurdish sentence does not match with another language. As shown in the below table:

کۆر مکه رۆیشت بۆ بازار.		The boy went to the market		
Parts of Speech				
1	کۆر	Subject	The	Article definite
2	مکه	Article definite	Boy	Subject
3	رۆیشت	Verb (past)	Went	Verb (past)
4	بۆ	Preposition	To	Preposition
5	بازار	Indirect object	The	Article definite
6			market	Indirect object

Furthermore, Izadin (2005: 94) shows that in English language, the inflectional suffix of plurality (~S) when adds to a noun will not change the part of speech while for Kurdish (~ان) has many functions and considers as derivational suffix, since it changes the part of speech, meaning and intention of the word in the sentence.

Making proper nouns like (سهیران، هۆزان، ریزان)

Making organization names (مامۆستایان، قوتابیان)

Making place names (چۆمان، حاجی ئومهران، شیخان)

As a symbol of duration, (گۆلان، گه‌لاریزان)

Showing pleasure and joy (خه‌نه‌ به‌ندان، خه‌ته‌نه‌ سوران)

Showing coldness (سه‌هۆل به‌ندان)

Showing homonym (سه‌مران) means head of tribes and someone in high positions in the society or community.

There is a verb in Kurdish language which is problematic for machine translation, since there is no auxiliary and modal verbs in Kurdish language except in a case of present which is (ه) and past (بوو). This (ه), in order to function as a verb, it should be added to an adjective. The problem is that there are some cases in the language that the (ه) loses its function as a verb and participates in changing the part of speech with different grammatical functions. As in:

کوشته = ه + و. v. کوشته (adj.)

سه‌موزه = ه + adj. سه‌موز (n.)

کۆره = ه + n. کۆر (n.)

It can be differentiated through their use in the sentence. (ibid: 75, 126). Apart from the above examples, Nariman (2010: 120-122) shows some additional functions of (ه) in Kurdish language. Such as:

1. Verb: شیلان زیره‌که. Shilan is clever.

2. Infix to make compound noun: گۆل+ه+هیرۆ = گۆله‌هیرۆ

3. Pronoun attaching a verb referring to second person singular and becomes subject: بخوینه.

4. Preposition: We went to the picnic. چووینه سمیران.

For zare (2005: 98), Sometimes in Kurdish language the agreement between the noun (subject) and the verb according to the form (not function) distracts or breaks as in;

(plu.) (Singular) مندالی همژار همن. (plu.) There are a lot of poor children (plu.)

Aurahman (1979: 13-14) believes that as a result of developing Kurdish language, parts of speech change, the elements of a sentence may change its function and take a new one. For example in Kurdish language

ههمن is (adj.) in: (Azad is a calm child.) - ئازاد مندالیکی ههمنه.

ههمن is (sbj.) in: (Hemn is a clever boy.) - ههمن کورپکی زیرهکه.

ههمن is (adv.) in: (Dara walked calmly.) دارا ههمن دهرۆیشت.

In this case ههمن can function as a (sbj., adj., and adv.)

Concerning pronouns, in Kurdish comparing to English, they rise another problem for MT, since Kurdish pronouns are bound (attaching the verb) and free morphemes while English pronouns are only free. As shown in table below by Omed (2011: 33)

number	Person	English free morpheme	Kurdish free morpheme	Kurdish bound morpheme	
				Transitive verbs	Intransitive verbs
1	singular	I	من	م	م
2		You	تۆ	ت	ت/یت
3		He, She, It	ئهو	ه	ه/یت/ات, Ø
1	Plural	We	ئێمه	مان	مین
2		You	ئێوه	تان	ن
3		They	ئەوان	یان	ن

Differences in Tenses

Tenses that exist in English language may not exist in Kurdish language. For example, in Kurdish language there is no form for differentiating present simple, present continuous, and future simple. As in :

ئهو دئ. = S/he comes.

S/he is coming.

S/he will come. (Salam, No year: 97)

Semantic Issues

Semantics is about studying and investigating the direct meaning in language (Huford et al, 2007:1). In MT the semantic problem come to pass if there would be lack in encoding items. Talib (2014: 44-47) also states that in modern linguistics, the problem between grammar and denotation words are differentiated according to ungrammatical sentences and non-semantic words. For example, it is easy to figure out the ungrammatical sentence as in:

* مندا لهکان له ژوور هکهدا نووست.

(This sentence is ungrammatical since the verb نووست is used with single pronoun and the subject is plural)

Also in * دار سیوهکه زیرهکه. * The apple tree is clever.

* سیروان کتیبهکهی خوارد. * Sirwan ate the book.

The above sentences are grammatically correct while semantically not. Since there is no agreement between the elements of the sentence, Hence, for this case the (syntagmatic and paradigmatic) level of sentences should be considered in translation. The words that go with each other should be highlighted, if not the syntactic rule refuses them and regards them as ungrammatical or incorrect. However; in some cases, those sentences can be meaningful and expressed correctly if they are in the negative or interrogative forms;

English	Kurdish
Who says Sirwan ate the book?	کێ دهڵێ سیروان کتیبهکهی خوارد؟
Sirwan didn't eat the book.	سیروان کتیبهکهی نهخوارد.

Pragmatic Issues

Another challenge for MT is that not only vocabularies and rules are sufficient for a good translation, but also past experiences have an important role. Pragmatically it is considered that the meaning of any word depends on the context and shared knowledge. As stated by Yule (2006:248) pragmatics is “the study of speaker meaning and how more is communicated than is said”.

Aurahman (2002: 135-136) remarks that arranging and differentiation among homonym, polysemy, and synonym in making dictionaries are really hard job and need an accurate intention. It is also worth mentioning that many Kurdish dictionaries still do not tackle this area. To him, a good way for making Kurdish dictionary is that the words should be explained briefly as how to use them. So in this case, many of the words in the dictionary should be put in simple sentences in order to clarify the word more. Likewise, Avesta (2009: 92) shows that Implicature, in any language, is a challenge pragmatically, and Kurdish language is not out of the situation.

A. پارهی پسوولمهکت دا؟ (Did you pay the bill?)

B. ئەمەین ئەمرو رۆژیکی خوش ئەبێت. (They say, today will be a nice day.)

Here, it shows that (A) intends to fight, (B) purposely says (today will be a nice day) and the meaning is controversial. Talib (2014: 61-62) indicates another challenge in Kurdish language for MT, since the meaning and intention of a sentence vary by changing the tone i.e putting pauses for sentence is to show; criticize, humor, disagreement, anger, surprise...etc. For example;

کۆریکی زیرهکه. (He is a clever boy.)

In the above sentence the speaker can use a (low-high) sound to show humiliation and criticize, and the actual meaning will be (he is not a clever boy). To extend that, Metaphor also in every language is also a problematic case for translation, every society has its own metaphor to compare and express ideas in a sentence, and Kurdish linguists argue that in Kurdish language there are sentences that have only one intention and meaning while there are some that stands for more than one meaning. Kamil (1981: 85-86) supports that idea by providing an example:

له خمو همستن میللمتی کورد خمو زهرمر تانه.	Wake up Kurdish people, sleeping is unbeneficial for you.
---	---

Here, the metaphor is له خمو همستن = wake up, which means don not be silent, be aware, revolute...etc. In addition, there is another obstacle for MT that is a tough task to the machine to provide a suitable meaning in the right position. This problem is called "ambiguity". One Kurdish word may have many different English meanings or vice versa. There is ambiguity in a language according the word and sentence level. Talib (2014: 51, 65-71) explains the ambiguity in the syntactic level as:

گهلاویژ ههلهات.
له شیره که نهگه ریم.

To understand the ambiguity in syntactic level, the purpose of using words in those sentences should be considered. To show the exact meaning some other phrases should be added to the sentence since for Kurdish language in the first sentence (گهلاویژ) means (A proper noun/ A Kurdish month), (ههلهات) means (escape/ appear) and the second sentence in English means (I am searching for the milk/sward.) so the word (شیر) stands for (milk/ sward). Similarly, Nariman (2012: 22) provides a Kurdish sentence that has double meaning: (پی = leg/ ability)

پیم دشکی.	I have ability to break it.
	My leg is breaking.
	My leg will break.
	My leg breaks.

Also putting stress on different elements of same sentence may cause ambiguity, for example:

Kurdish	English
'همر چواریان هاتن.	The all four came.
همر 'چواریان هاتن .	Only four of them came.

Also sometimes the same sentence may contradict in meaning when it is translated. As in:

Kurdish	English
نهرۆی دهتکوژم.	If you don't go, I will kill you.
نهرۆی 'دهتکوژم	Don't go, or I will kill you (If you go, I will kill you).

A stop in a sentence may change the meaning of it. As in

Kurdish	English
خوشکه کهتان ژنی کامهیانه؟	The sister (Katan) is the wife of whom?
خوشکهکهتان ژنی کامهیانه؟	Your sister is the wife of whom?

In the above sentences, the first sentence is about someone whose name is (Katan = کهتان) which is a proper name, while in the second sentence the speaker or writer means (someone's sister i.e your sister = خوشکهکهتان) (Talib, ibid).

According to Kamal (2007: 123) ambiguity in the word level occurs when it (in a sentence, a phrase or a word) has double meaning. In Kurdish language, for this case, the expressions (without other meaning, to be far from you) use to save the situation and make the reader or hearer not to understand it in a wrong way. As shown the under lined words in the following sentences:

Kurdish	دکتۆر به نمووشمکهی ووت: هانی، دهتوانی قووتی بدهی، دهشتوانی بیمری.
English	The doctor told the patient: here it is, you can swallow it, and you can also <u>suck</u> it.

3. Other Obstacles and Issues

Political

There are some complications for MT. the most obvious one is the lack of unified language for all Kurdish dialects, which can be considered as one of the main issues with Kurdish MT. Neither Sorani nor Kurmanji dialect speaker can get benefit from “google translator” nor “inkurdish translator” since the first one is designed for Kurmanji and second one for Sorani. The challenge with this issue is out of researcher's control because this is the responsibility of Kurdish political authority. As Izadin (2005:119-120, 156) clarifies that in Baghdad Radio (1958) the program published in Kurdish language but in two different dialects. In addition, in 1956, a national journal published that the directors and writers didn't understand the intended purpose of that journal, and the publishing discourses were in two different dialects of Kurdish language. Unfortunately, until now this procedure is used and depended on in journals and newspapers. This shows the idea as if Kurdish language has two dialects rather than a standard language. For this reason, many believes that Kurdish language is not only one language and occupiers hope to differentiate between the dialects under the realm of not understanding each other, and ask for publishing two types of books in which this demand leads to a great problem of deviating the language. In his book, Hussain (2006: 18-22) explains that the Islamic political parties also affect the idea of not having a standard Kurdish language, since

in their activities; they are using many Arabic words while talking in Kurdish. They also focus on Arabic culture, language, history, and using many Arabic words, in giving lectures or talking in public, rather than focusing on Kurdish language and culture. Moreover, He says that there is a serious danger on Kurdish language since Kurdish names and greetings are about to be replaced and become Arabic. Apart from that, Kurdish society is occupied by Turkish and Persian government in which again the shared feature among them is the Islamic religion, thus they prevent Kurdish people to speak Kurdish and force them to use other languages like (Arabic, Persian, and Turkish). So, geographically occupiers divide Kurdish people in to four locations (Turkey, Iran, Syria, and Iraq). The Turkish government considered Kurdish people as a (Turkish ethnic groups living in mountains), and nowadays, they state that there are Kurds but no Kurdistan state and they will not consider Kurds as having different society and culture. In Iran, the Persians do not allow Kurds as being and having a separate and independent society. Yet, the Arabs in Iraq and Syria also regard Kurdish people as minority without land, since geographically both countries are part of Arab Union Nations. Furthermore, ethnically the population of both countries is regarded as part Arab nation.

Social

Omer (2015: 57) states “there are social norms that differs from one society to another”. For instance, (dog) in Kurdish society is considered as something disgusting and negative while in English society it is positive and it represents a symbol of loyalty. Considering the following examples to show that difference:

Kurdish	دهریا به دهمی سمگ پیس نابیت.	The river will not be dirty by a dog's mouth.
English	Work like a dog.	In Kurdish society means work obligatory and by force

Hence, it can be said that society and culture also effect on the meaning. For example, in Kurdish language (خړ= round) is a symbol of beauty for appearance of a female, and (لاکێشه=rectangle) stands for ugliness. For example;

A: كچهكه چۆنه؟ (How does she look?)

B: - خڕه. (Means she is beautiful)

- لاکێشهیه. (Means not beautiful)

Sometimes the shared knowledge plays a great role in understanding or expressing an idea. For example, in Kurdish society when someone wants to praise someone else he/she says:

He is a Mama Risha.

(Mama Risha) is a Kurdish famous personality for his braveness, honesty, and being a good fighter. If one doesn't have knowledge about who Mama Risha is, s/he may miss the real meaning of that sentence.

4. Conclusions

1. As a first step, it is better to work on unidirectional translation (English to Kurdish) instead of bilingual translation (English to Kurdish and Kurdish to Kurdish).
2. Deep analysis of Kurdish language compatibility for machine translation is necessary.
3. Translation between languages from semi source (Indo-European) such as Kurdish and Persian language has more quality output than translation between languages from totally different sources such as Kurdish and English.
4. Many issues are still found regarding MT with not being solved yet, and the main issue would be that of regarding standard language for Kurdish.
5. Kurdish language comparing to the English one is a difficult issue for MT, since Kurdish is an agglutinative language, in which a sentence might be expressed with a word, and the elements of a sentence become bound morpheme.

References

- ئهورهحماني حاجي مالف، ١٩٧٩، ريزماني كوردى، بهغدا: چاپخانهى كۆرى زانيارى عيراق.
- ئهورهحماني حاجي مالف، ٢٠٠٢، بهر ههمه زمانهوانيهكانم، سلیمانى: وهزارهتى روشنبيرى.
- ئاقبيستا كهمال محمود، ٢٠٠٩، پراگماتيكي رستهى پرسيارو فهومان له زمانى كورديدا، سلیمانى: مملبهندی كوردۆلۆجى.
- ئوميد بهرزان برزو، ٢٠١١، دروستهى فريز و رسته له زمانى كورديدا، سلیمانى: چاپخانهى گنج.
- بابارسول نورى رسول، ٢٠١٢، پيکهاتهى ژماره له زمانى كورديدا، ههولير: چاپخانهى حاجي هاشم.
- تاليب حوسين عهلى، ٢٠١٤، ههندى لايهن له پهيوهندی نيوان رسته و واتا له كورديدا، ههولير: چاپخانهى حاجي هاشم.
- حوسين محممهدهم عزيز، ٢٠٠٦، ههندى بابتهى زمانهوانى و ويژهيى، سلیمانى: چاپخانهى سيما.
- زهري يوسوفپوفا، ٢٠٠٥، شيوهى سلیمانى زمانى كوردى، ههولير: چاپخانهى وهزارهتى پهروهرده.
- سهلام ناوخوش بهكر، نيه، زمانناسى و ههندى بابتهى زمانناسى كوردى، ههولير: چاپخانهى سوران.
- ساجيده عهبدووللا فههادى، ٢٠١٣، ههندى لايهنى رستهسازى زمانى كوردى، ههولير: چاپخانهى حاجي هاشم.
- عزهدين مستهفا رهسول، ٢٠٠٥، بو زمان، سلیمانى: چاپخانهى شقان.
- عومهر فهتحووللا وههاب، ٢٠١٥، پيوازوكاريى خوکاربيانهى زمانى كوردى، (ماستهنامه)، سلیمانى: زانکۆى سلیمانى
- كهمال ميراوهدلى، ٢٠٠٧، فههنگى ريزماني كوردى، سلیمانى: مملبهندی كوردۆلۆجى.
- كامل حسن عزيز البصير، ١٩٨١، زانستى ئاولواتا، بهغدا: چاپخانهى كۆرى زانيارى عيراق.

- نهریمان عهبدوڵڵا خۆشناو، ۲۰۱۰، ریزمانی کوردی، ههولێر: چاپخانهی مناره.
- نهریمان عهبدوڵڵا خۆشناو، ۲۰۱۲، رستهسازی، ههولێر: چاپخانهی رۆژههلات.
- وریا عومەر ئهمین، ۲۰۱۱، کهرهستهکانی دهرشتنی کار له زمانی کوردی، ههولێر: چاپخانهی حاجی هاشم.
- وریا عومەر ئهمین، ۲۰۱۱، ب، پیتۆکهکانی زمانهوانی، ههولێر: چاپخانهی ئاراس.

Crystal, D. (2003) *A Dictionary of Linguistics & Phonetics*. 5th Ed. Library of Congress Cataloging: Blackwell Publishing

Huford, R.J., Heasley, B., and Smith, B.,M. (2007) *SEMANTICS: a course book*. 2nd Ed. Cambridge: Cambridge University Press.

Leech, G., Cruickshank, B., and Ivanic, R. (2001) *A-Z of English Grammar & Usage*. England: Longman Publishing.

Quirk, R. and Greenbaum, S. (1973) *A University Grammar of English*. Hong Kong: Longman Publishing.

Yule, G. (2003) *The Study of Language*. 3rd Ed. Cambridge: Cambridge University Press.