

# Kurdish Sorani Dialect Morphology Generation Using a Concatenative Strategy



Kardo O. Aziz<sup>1</sup>, Ramyar A. Teimoor<sup>2</sup>, Tofiq A. Tofiq<sup>2</sup>, Dilman S. Abdulla<sup>3</sup>

<sup>1</sup>Department of Computer, College of Science, Chamro University, Iraq, <sup>2</sup>Department of Computer, College of Science, University of Sulaimani, Sulaimanyah, Kurdistan Region, Iraq, <sup>3</sup>Department of Computer, College of Science, University of Halabja, Halabja, Kurdistan Region, Iraq

## ABSTRACT

In natural language processing, morphological generation refers to the creation of the appropriate inflected forms of words based on a predetermined set of morphological rules. However, it might be difficult to generate morphology in languages with intricate morphological systems, like the Kurdish Sorani dialect. The concatenative morphology-based unique technique to morphological generation in Kurdish Sorani is proposed in this research. The suggested strategy tries to get over the drawbacks of current approaches and enhance the precision and effectiveness of morphological generation in Kurdish Sorani. The suggested technique generates all conceivable subjective and objective pronouns in both positive and negative forms, together with the various verb tenses for Kurdish morphology. The study presents a detailed examination of Kurdish Sorani's morphology and points out the difficulties in coming up with the right verbforms. The authors suggest a concatenative morphology-based morphological generating system that comprises of a morphological analyzer and a morphological generator.

**Index Terms:** Low-resource-language, Kurdish-Sorani, NLP, Word-Generation, Morphology

## 1. INTRODUCTION

In the digital age, language technology is a growing subject that depends on our understanding of human language and computer techniques to handle it [1]. Even though for some languages, such as English, the field of text processing has been studied very well and many significant efforts can be seen, but for the Kurdish language, despite of having a great number of speakers and computer users, the amount of research in the subject of Kurdish text processing is still rather little [2]–[4].

The Kurdish language is a part of the Indo-Iranian family of Indo-European languages. It is a language spoken by Kurds

and Kurdish people in Kurdistan (also known as “the region of the Kurds”). The most closely related and more widely used language to Kurdish is Persian. There is not a particularly exact statistical survey for the number of native Kurdish speakers in the globe, although it is believed to be more than 45 million [5]. It has diverse dialects. Kurmanji and Sorani are the most frequently spoken dialects in terms of both speakers and degree of standardization [3]. It has limited resources, and the relevant research, information, and tools are still in their infancy [6].

Morphological generation is the process of generating different word forms based on their grammatical properties. In computational linguistics, morphological generation is an important part of natural language processing (NLP), as it enables machines to generate grammatically correct and natural-sounding text. There have been several approaches to morphological generation, including rule-based [4], [7], stochastic, and hybrid methods.

In the Kurdish language, the sheer number of changes that must be developed for Kurdish and Persian verbal

### Access this article online

DOI: 10.21928/uhdjst.v8n1y2024.pp13-19

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2024 Aziz, *et al.* 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:** Ramyar A. Teimoor, Department of Computer, College of Science, University of Sulaimani, Sulaimanyah, Kurdistan Region, Iraq. [ramyar.teimoor@univsul.edu.iq](mailto:ramyar.teimoor@univsul.edu.iq)

Received: 14-10-2023

Accepted: 02-01-2024

Published: 10-01-2024

morphology is the main problem. Verbal stems have trilateral or quadrilateral origins (3 or 4 radicals). The conventional combination of root morphemes and vowel melodies forms the derivational foundation for stem construction. Roots, patterns, and stem production are referred to as interdigitation [4], [8].

It is a difficult undertaking to build NLP technology for Arabic and Kurdish. These languages have significant morphology and both derivational and inflectional forms [4], [9]. Furthermore, the complexity of Arabic and Kurdish morphology makes it challenging to analyze or create as a vital component of the whole simulation process. On the one hand, there are now many analysts working in the area of analyzing Kurdish morphemes, which has advanced considerably. These attempts can be seen in [1], [2], [6], [10]. These analyzers provide a variety of outcomes using various techniques. Despite their well-known limitations, they provide a trustworthy source for computational analysis, and in certain situations, even their shortcomings are seen to be beneficial. On the other hand, because of its widespread use in essential everyday applications like spell checking, word generation has also drawn a lot of interest [10]–[13].

In this work, we provide a computational technique that applies a special treatment to the creation of Kurdish morphology that is described in the methodology parts by separating the issue of infixation from other inflectional variations. Since short vowels and diacritical marks are pedagogically advantageous and can always be deleted before presentation, we design Kurdish verbs with them.

The remaining components of this study are listed below. In Section 2, we review the literature on morphological analyzers and stemmers in Kurdish, Arabic, and Persian Language. in Section 3. To assess the effectiveness of the model, categorization results are compared in Section with total accuracy in the fourth section, and the conclusions are in the fifth section.

## 2. RELATED WORK

There are other works that are connected to this effort, but the most of them are stemmers and word generators for the Arabic language and Persian. We looked at their methodology and developed a new one for our Kurdish Sorani dialect. Most of the related studies work on analyzing morphological tokens and attempt to eliminate the suffixes and prefixes,

while our work is the inverse of their work. We provide the root or the stem and try to append prefixes and suffixes to generate all variations of a single verb.

Ahmadi [4] provides a thorough description of Kurdish-Sorani morphological and morphological constructions in a formal way to tackle complex morphology of Sorani. Ahmadi has presented morphological word generation rules with most of the possibilities and provides a couple examples for each tense. Furthermore, Ahmadi claims that morphemes often have different phonological forms depending on the structure of the word and the surrounding morphemes. Sorani Kurdish clitics and affixes often exhibit allomorphy, that is having more than one single shape for the same morpheme. Ahmadi has done an extensive study on Kurdish Sorani Morphology, but since the study is just in theory and descriptive. Thus, they do not provide a program or system for practical and evaluation or for production. Thus, we exploit this issue and we reuse some of their written rules that is related to Kurdish verbs along with some improvements to create an automatic system to generate words from a list of Kurdish verbs

Naserzade *et al.* [14] offer a thorough morphological analyzer for Central Kurdish (CK) in this study. They initially put together and methodically classified a complete collection of the language's morphological and morphophonological rules, building on the few available literature. A generative lexicon with almost 10,000 verbs, nouns, adjectives, named entity, and other forms of word stem stems was also compiled and carefully categorized. They implemented the CKMorph Analyzer based on finite-state transducers using these rule sets and resources. They gathered test sets for assessing the analyzer's correctness and coverage, personally labeled them, and made them available for public use as a baseline for future study. About 95.9% of the 1000 CK words in the first test set, which CKMorph assessed morphologically and contextually, were properly analyzed. Furthermore, of the 4.22M CK tokens in the second test set, 95.5% were subject to at least one analysis from CKMorph. Their work is almost the inverse of our work. While we work on generating morphological words, specifically verbs, they analyze those morphological words and label each part or section of each word.

Another study which is Saeed *et al.* [15], they study morphological tokens in Kurdish and remove suffixes and prefixes to get the original word or verb that created the word on. Thus, the work uses pre-created tokens which have many prefixes and suffixes. The goal of this study is to develop a method for classifying Kurdish using Reber Stemmer. As a result, a novel method is being researched to get the stem

of Kurdish words by eliminating their longest suffixes and prefixes. This method has a strong capacity and satisfies the criteria when it comes to the process of getting the stem of Kurdish words by removing as many of the necessary affixes as feasible. This stemmer's benefit is that it disregards the list of affixes in the right stem ordering for multiple words with the same format. The eight-class KDC-4007 dataset is used to apply the stemming approach. The categorization is performed using the Support Vector Machine (SVM) and Decision Tree (DT or C 4.5). The Longest-Match stemmer approach has been effectively compared with this stemmer. The F-measure of the Reber stemmer and Longest-Match methods in SVM is greater than DT, the data show. Reber stemmer in SVM achieved higher F-measure for the classes (religion, sport, health, and education), whereas the remainder of the classes are lower in Longest-Match. Reber stemmer had a greater F-measure in DT for the courses (religion, sport, and art), but in the longest match, the rest of the classes had a lower F-measure. The intuition of our study came from this and similar papers. We looked into their dataset and found that most of the datasets that are used for stemming do not cover all the possible morphological words that could be created from a single stem. We believe that to trust a stemmer in real life, it should understand all the possible morphological words. We found that most of the missing cases were those morphological that appear least or not appear at all in the dataset.

Furthermore, Yoosofan *et al.* [16] state that the basic problem with generating non-concatenative languages such as Arabic verbal morphology is the large number of variants that must be generated. They compare Arabic and Persian Stemmers and the features in each language. One of the challenges for stemmers are the enormous variations of a single word.

Each word must be recognized and correctly stemmed by a stemmer. Regarding Arabic, the evolution and variations of these words provide the fundamental challenge to their derivation. When compared to Persian terms, Arabic words exhibit distinct derivational behavior morphologically. In addition, several of these Persian terms have distinctive qualities that allow us to separate them from Arabic counterparts. They have limited their derivation to a few common trilateral roots to get the right results. For Example, Information retrieval, text classification, text summarization, automated phrasal category recognition, translation studies, NLP, etc.

A very similar work which has done the same work as ours but for Arabic is Aqel *et al.* [11]. They have developed a method that will deliver almost all the words that can be

produced out of any submitted word. Their study will address the issue of whether or not derived and inflected words can both be formed using the same process. Furthermore, since morphology analyzes the word structure taking into consideration its fundamental meaningful components, several suggestions for building the method described here are considered. It has consistently been one of the most crucial elements in almost every application of NLP. That idea was applied to Arabic, which resulted in heavily inflected and derived terms. In their work, a method is devised that yields nearly all of the words that may be generated from any given word.

TMT Sembok and Ata [17] show the importance of a good stemmer in information retrieval tools. They showed that a stemmer increases the effectiveness of retrieval systems. The results have demonstrated that retrieval effectiveness has increased when stemming is implemented because document retrieval in information retrieval systems (IRS) often comprises acquiring relevant documents related to information needs. The system's ability to understand document content will boost the efficacy of the retrieval results. However, it takes a lot of work to understand the subject. Both stemmed and unstamped keywords are available. When they stem and conflate keywords during the retrieval process, they improve the usage of semantic technology in IRS. In morphological analysis, which occurs before syntactic and semantic analysis in NLP, word stemming is a step.

In addition, Habash and Rambow [18] provide MAGEAD, a morphological analyzer and generator for Arabic language. Their research is groundbreaking since it specifically addresses the necessity for processing the languages' morphology. MAGEAD combines morphemes from several dialects and conducts an online analysis to or generation from a root+pattern+features representation. It also contains independent phonological and orthographic representations. They give MAGEAD a thorough analysis. MAGEAD has a very similar work to our study. They introduce a number of rules and a list of prefixes and suffixes that will possibly attach to a stem to generate morphological tokens.

### 3. METHODS

We started by collecting the verbs that we have in Kurdish language from academic and non-academic documents. Generating verbs for all of the tenses and persons including negative and positive tenses are all relays on the verbs' (base, and root). Verbs are chosen for this paper because the

Kurdish language has verbal incorporation features. Incorporation is a morphological process in which a verb is combined with another word, usually a noun or an adjective, to form a complex word that expresses a complete idea or sentence. In Kurdish language, as well as in other languages that use incorporation, the resulting word can act as a standalone sentence without the need for additional words or elements. For example, the word (“ناتاندەمی”/nætəndəmə/) means “I will not give it to you” in English and a whole sentence was required to express the same meaning. It contains the “I” and “you” pronouns, which can be replaced by other personal pronouns to change their meaning. In addition, the negative word is also included which can be removed to get a positive form of the word. Furthermore, prefixes, suffixes, or pronouns cannot be simply replaced or removed to change their meaning. In most cases, the entire sequence of prefixes, suffixes, and stemming is changed.

We have three types of verbs in Kurdish. Simple verbs have a single word and specific meaning (e.g., بردن/brdn/means “taking”). Compound verbs are created by combining an adjective or a noun with a simple verb (e.g., نارامگرتن/ər'ɑ:mgrtn/means being patient). Finally, Complex verbs, which created by add a prefix, or suffix to a simple verb [7], (e.g., تێکدان/tēkdən/means spoiling created by a prefix (no-meaning) + simple verb, بردنەو/brdnəw/created by a simple verb + a suffix (هه) means to win).

For each of these verbs, we manually found their bases (i.e., past stem) and roots (i.e., present roots), because each verb has a different structure it cannot be done by a program. All Kurdish verbs(Chawg) end with “ن/n/” letter hence, bases are created by dropping the “ن/n/” letter at the end. We replace it by an underscore. Regarding their roots, most of the verbs have different roots for past and present depending on the verb. Thus, we have multiple roots for each verb. Sometimes the verb has a root which is very far from the verb in terms of writings. We place an underscore in the places which have the possibility of adding prefix or suffix or affix.

Examples of bases are shown in Table 1.

Examples of roots are shown in Table 2.

These underscores will be replaced by pronouns and/or negation words or a special letter based on the tense and person you want to generate.

In Kurdish, the verbs(infinitive) are ends with letter “ن” as we mentioned before and before the “ن” letter one of these

**TABLE 1: bases example**

Verb	Base (Past stem)
بردن	برـ
نارامگرتن	نارامگرـ
تێکدان	تێکـ

**TABLE 2: Root Example**

Verb	Root1	Root2
بردن	برـ	برـ
نارامگرتن	نارامگرـ	نارامگرـ
تێکدان	تێکـ	تێکـ

five letters are coming (“د/d/,” “ی/i/,” “ت/t/,” “و/w/,” “ئ/ə/”). Most of the times, these letters have effect on which modal prefix/postfixes will be used. For example, if it has “و” letter before the “ن” we cannot use “وو/u:ə/” because will make the generated word have three consecutive “و” letter which is not correct in Kurdish.

### 3.1. Verb Types

The verbs are divided into two types (intransitive and transitive). Intransitive verb is a verb that cannot take an object (e.g., هاتن means coming). Transitive verbs take an object (eg. شستن means washing). Each type has different rules for word generation and takes a different set of bound pronouns (BP) which is shown in Table 3. Each word is manually marked which set of BP accepts.

### 3.2. Tenses

Future sense is gained from context or by time expressions. While the present tense is formed by adding one of the modal prefix such as “ده” or “ته” before the root of the verb and a subjective bounding pronoun at the end (modal prefix + root + Subjective bound pronoun) example, (بردن, root: برـ present simple: دهبرم. While “ده” is the modal prefix and “م” is bound pronoun for me).

There are many past tenses in Kurdish language. All of them can be formed by adding subjective BP and/or modal prefixes.

### 3.3. Subjective BP

Pronouns for Past tenses Past tenses are using the following subject pronouns set for transitive and intransitive verbs, see Table 3.

Pronouns for Present tenses

Present tenses are using the following subjective bound set for both transitive and intransitive verbs, see Table 4.



**TABLE 3: Past bound pronouns**

Person Type	Transitive		Intransitive	
	Singular	Plural	Singular	Plural
1 <sup>st</sup> Person	م	مان	م	ین
2 <sup>nd</sup> person	ت	تان	یت	ن
3 <sup>rd</sup> person	ی	نای	None	ن

**TABLE 4: Singular and plural example**

Person Type	Singular	Plural
1 <sup>st</sup> Person	م	ین
2 <sup>nd</sup> person	ی/یت	ن
3 <sup>rd</sup> person	ئ/ئیت	ن

Negative tenses are created by adding one of these two modal prefixed (“نە”، “نا”) with wither past stem or roots [7]. The position of the prefix is varying based on the tense you want to generate. It will be mentioned in the Rules section.

### 3.4. Rules

Rules of generating verbs in Kurdish vary based on verb type, tense, and negativity, thus we have to check every verb we face and specify the properties and most importantly which tense we are generating.

For most of the tenses, we have a special modal prefix and/or suffix or postfix to represent that tense. For each verb type the position of the modal prefixed is different for example, In the simple verbs, the prefixes are going to the beginning of the past stem but in complex verbs the prefix goes after the first preposition. In compound verbs, the prefix goes after a pre-word which is usually a noun or an adjective attached to a verb to create the compound verb. Sometimes the verb has more than one pre-word or preposition, thus some of the prefixes will be placed after the first pre-word and some of them (such as negative words) placed after the last preposition.

We have written approximately 94 rules to create the words. Simple verbs have 12 rules, complex verbs have 36 rules. compound verbs have 34 rules. Finally, we have 12 rules for present tenses for all types of the verbs. We will not be able to write and explain each of the rules. We have shown some of the rules for each verb type below.

### 3.5. Simple Verbs

Simple past:  $[Past\ stem][Subjective\ BP]$  (eg. بردم/brdm/ means I took)

Negative simple past:  $[Negative\ word][Subjective\ BP][past\ stem]$  (eg. نەمبەرد/nəmbərd/ means I did not take)

Past continues:  $[Modal\ Prefix][Subjective\ BP][Past\ stem]$  (eg. دەمانبرد/dəmənbərd/ means we were taking)

### 3.6. Complex Verbs

Simple past (transitive):  $[1^{st}\ prefix][Subjective\ BP][other\ prefixes][past\ stem]$  (eg. رابەینا/rəməbēnə/ means I brought it up)

Simple past (intransitive):  $[all\ prefixes][past\ stem][Subjective\ BP]$  (eg. رابەاتم/rəhətm/ means I'm used to it)

Negative simple past (transitive):  $[1^{st}\ prefix][Subjective\ BP][other\ prefix][negative\ word][past\ stem]$  (eg. رانەهینا/rənəhēnə/ I did not bring it up).

### 3.7. Compound Verbs

Past continues (transitive):  $[1^{st}\ word\ of\ past\ stem][Subjective\ BP][negative\ word\ if\ negative][modal\ prefix\ “دە”/nə/\ or\ “ئە”/ə/][rest\ of\ the\ past\ stem]$  (eg. بەکارماندەهینا/bəkərmnəbēnə/ means We were using it)

past perfect (transitive):  $[1^{st}\ word\ of\ past\ stem][Subjective\ BP][negative\ word\ if\ negative][rest\ of\ the\ root][modal\ postfix\ “وە”/wə/\ or\ “ووە”/u:ə/]$  (eg. بەکارمەیناوە/bəkərmhənəwə/ means I have used it)

Care must be taken when a modal prefix is added to a verb, as they may generate invalid character sequences. For example, it is possible to make three consecutive identical characters appear when a specific modal prefix is added to a specific verb. In this case, one of these characters must be removed as three consecutive identical characters in Kurdish Sorani are not allowed.

Since we have identical subjective BP in the tables for past and present tenses, sometimes we get the same word for two different people. Just like in English “you” is used for second person singular and plural. To differentiate between them we need to put it into a sentence. We have a similar pronoun which is the letter “ن” which is used in the second person plural and third person plural. Thus, the number of generated words is not equal for all of the verbs.

There is another issue while generating words, which is having two consecutive vowel letters. In this case, based on Kurdish grammar rules and depending on which vowel letter is, we had to add an extra character in between the two consecutive characters.

Furthermore, the transitive verbs also accept pronouns as objects. Thus, possibly most of the time the verbs have the

pronoun attached to them. It means that it will make the process more difficult to figure out where to put the pronoun that acts as object. This will also make the rules more complicated. For example, if you want to say “They did not take me” in English, it will be “نهیانبردم/nehjænbrdm/” in Kurdish. The “نه” in the beginning is the negation word. The “یان” that comes after means “they”. Then, the past stem comes. Finally, the “م” at the end of the word means “me”. The position of each of these words will vary a lot depending on the verb type, tense, and negation especially in compound and complex words. For example, if we remove the negation word “they took me” it will be “بردیانم/brdiæn/m/” in Kurdish. Note that the “یان/iæn/” pronoun went after the past stem before the subject pronoun of “م/m/”. More examples are shown in Table 5.

#### 4. RESULTS AND CONTRIBUTION

An enormous number of Kurdish verbs (infinitive) are collected and marked with their properties, such as verb types (simple, complex, and compound). The verbs past stem, past root, and present roots are specified. The position of pronouns and negative words is marked by replacing an underscore. This underscore will help our next steps to find where exactly each modal prefix or pronoun will be placed. Each verb is verified by a Kurdish language specialist to make sure they are available and correct. Total number of verbs is 2463. The number of verbs based on their type is shown in Table 6.

**TABLE 5: Present bound pronouns**

Word	Negation word	Negation word removed
نهتدمکۆلی	نه	دتدمکۆلی
نهتاندەتاشین	نه	دتاندەتاشین
نایاندەبینەوه	نا	دیاندەبینەوه

Some examples of the verbs are shown in Table 7.

**TABLE 6: Verbs example based on their types**

Verb type	Transitive	Intransitive	Total
Simple	181	175	356
Compound	1362	247	1609
Complex	357	141	498
Total	1900	563	2463

Then, a list of rules is defined. These rules cover all the possible tenses and subject persons with both positive and negative tenses. Using these rules, a list of words for each verb is generated in Fig. 1. We have generated approximately 317 words per verb. These words represent the verb for a specific subject person and possible specific object person in different tenses. These words are all the possible appearances of the verb in any text. We have gathered around 781,862 words in total including verb gerunds and negatives.

**TABLE 7: Generated word examples**

Verb	Some examples of generated words
/dru: stkrdn/	درۆستکردن، درۆستکارێ، درۆستکارێت، درۆستکارێت، ...
/trsæn/ ترسان	دەترسام، نەدەترسای، نەدەترسان، نەترسیت، ترسای، ترسای، ترسای، ...

These words will be a very helpful corpus for the next research project that works on NLP on Kurdish language. Furthermore, it will help with Kurdish spell error detection and checking systems. Researchers can have a ready to go dataset that includes all the Kurdish verbs in all different tenses and persons including negative tenses. Especially all the researches and papers on stemming can have a great dataset that can verify and train their model on. It can also be a great state-of-art dataset to compare their results on.

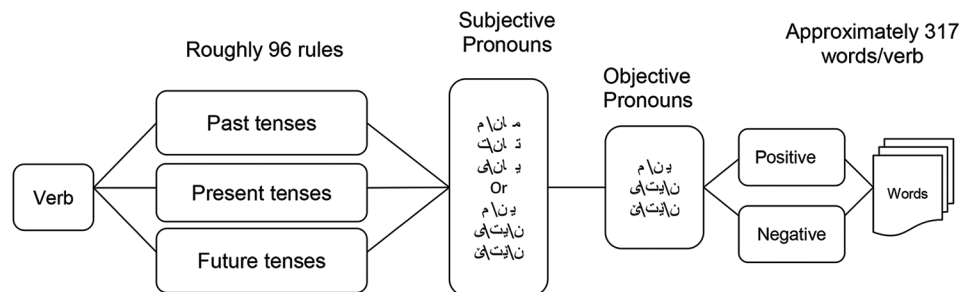


Fig. 1. The diagram of the process for the word generation.

List of the verbs that are generated in our work and the rules that we have used are available from this link in Excel format along with, all the words in this GitHub repository [19].

## 5. CONCLUSION

Our approach was motivated by helpful factors. We explored efficient methods to use a morphological generating tool, which is a common setting for us to construct automated translation systems and a stemmer. The number of rules for morphological transformation that must be followed. In this study, a technique that can create almost all the words from any input word was devised. After the word has been thoroughly examined, it will first be submitted, and its characteristics will be listed. Accordingly, depending on the appropriate affixes to the word supplied from the user, new inflected words are created utilizing the approach established and illustrated through this study. This study also addresses the issue, “Can inflected and derived words be created equally using the same methodology?” It was shown that, to do that, a different model based on the Root, Base and Patterns is needed as opposed to the one constructed based on the Stem and Affixes. We managed to build a large number of rules with specific suffixes and prefixes that can cover every tense, subject, and object pronouns with positive and negative states. We achieved 781,862 fully verified words that can be used for stemmers and translated. The word is publicly available for other researchers in an Excel file that mentioned in the end of the result section. For future research directions, we acknowledge that our algorithm exhibits certain complexities and could benefit from enhanced code reusability. We invite interested researchers to explore methods for reducing the number of rules and optimizing the reuse of existing ones to improve the algorithm’s speed and efficiency. In addition, although we have made a comprehensive effort to include a vast array of Kurdish verbs in our work, there remains scope for further verification and expansion. Future researchers may contribute by identifying and integrating additional Kurdish verbs that may have been overlooked in our current dataset.

## REFERENCES

- [1] S. Ahmadi, “KLPT-Kurdish Language Processing Toolkit,” In: Proceedings of Second Workshop for NLP Open-Source Software (NLP-OSS), pp. 72-84.
- [2] S. Salavati and S. Ahmadi. “Building a lemmatizer and a spell-checker for Sorani Kurdish”. arXiv preprint, vol. 1809.10763 , p. 1, 2018.
- [3] D. Salih, “Kurdish Sorani Spelling Checker System,” [MA Thesis], University of Birmingham, England, 2016, 2021.
- [4] S. Ahmadi, “A formal description of Sorani Kurdish morphology”. ArXiv Preprint, vol. 2109.03942, p. 1, 2021.
- [5] F. I. Kurde de Paris. “The Kurdish Population”. 2017. Available from: <https://www.institutkurde.org/en/info/the-kurdish-population-1232551004> [Last accessed on 2023 Dec 24].
- [6] R. O. Abdulrahman and H. Hassani, “A Language Model for Spell Checking of Educational Texts in Kurdish (Sorani)”. In: Proceedings of the 1<sup>st</sup> Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages, SIGUL 2022-Held in Conjunction with the International Conference on Language Resources and Evaluation, pp. 189-198, 2022.
- [7] H. Fatah and Z. Hamawand. “A prototype approach to Kurdish prefixes”. *International Journal on Studies in English Language and Literature*, vol. 2, pp. 37-49, 2014.
- [8] V. Cavalli-Sforza, A. Soudi, and T. Mitamura. “Arabic Morphology Generation using a Concatenative Strategy”. In: 1<sup>st</sup> Meet. North American Chapter of the Association for Computational Linguistics. NAACL 2000-co-Located with 6<sup>th</sup> Applying Natural Language Processing Conference, pp. 86-93, 2000.
- [9] R. A. Kareem. “The Syntax of Verbal Inflection in Central Kurdish”. Newcastle University, England, 2016. (Doctoral Dissertation).
- [10] S. Ahmadi. “Hunspell for Sorani Kurdish spell checking and morphological analysis”. Arxiv, vol. 2109.06374, p. 1, 2021.
- [11] A. Aqel, S. Alwadei and M. Dahab. “Building an Arabic words generator”. *International Journal of Computer Applications*, vol. 112, pp. 36-41, 2015.
- [12] D. H. Kim. “A Basic Guide to Kurdish Grammar”. Culture and Language Institute of Kurdi and Kori, Iraq, 2010.
- [13] G. Walther. “Fitting into morphological structure: Accounting for Sorani Kurdish endoclitics”. *Mediterranean Morphology Meetings*, vol. 8, pp. 299-321, 2012.
- [14] M. Naserzade, A. Mahmudi, H. Veisi, H. Hosseini, M. MohammadAmini. “CKMorph: A comprehensive morphological analyzer for Central Kurdish”. *International Journal of Digital Humanities*, vol. ???, pp. 1-46, 2023.
- [15] A. M. Saeed, T. A. Rashid, A. M. Mustafa and A. A. Agha. “An evaluation of Reber stemmer with longest match stemmer technique in Kurdish Sorani text classification. *Iran Journal of Computer Science*, vol. 1, no. 2, pp. 99-107, 2018.
- [16] A. Yoosofan, A. Rahimi, M. Rastgoo and M. M. Mojiri. “Automatic stemming of some Arabic words used in persian through morphological analysis without a dictionary”. *World Applied Sciences Journal*, vol. 8, no. 9, pp. 1078-1085, 2010.
- [17] T. M. T. Sembok and B. A. Ata. “Arabic word stemming algorithms and retrieval effectiveness”. *Lecture Notes in Engineering and Computer Science*, vol. 3, pp. 1577–1582, 2013.
- [18] N. Habash and O. Rambow. “MAGEAD: A morphological analyzer and generator for the Arabic dialects.” In: COLING/ACL 2006-21<sup>st</sup> International Conference on Computational Linguistics. 44<sup>th</sup> Annual Meeting of the Association for Computational Linguistics. vol. 1, pp. 681-688, 2006.
- [19] K. O. Aziz. “Kurdish-Morphological-Kurdish-Word, Rules and Source Code”. 2023. Available from: <https://github.com/kardoothman/kurdish-morphological-kurdish-word> [Last accessed on 2023 Feb 01].