HUNSPELL for Sorani Kurdish Spell Checking and Morphological Analysis

Sina Ahmadi
Insight Centre for Data Analytics
National University of Ireland Galway
ahmadi.sina@outlook.com

Abstract

Spell checking and morphological analysis are two fundamental tasks in text and natural language processing and are addressed in the early stages of the development of language technology. Despite the previous efforts, there is no progress in open-source to create such tools for Sorani Kurdish, also known as Central Kurdish, as a less-resourced language. In this paper, we present our efforts in annotating a lexicon with morphosyntactic tags and also, extracting morphological rules of Sorani Kurdish to build a morphological analyzer, a stemmer and a spell-checking system using HUNSPELL. This implementation can be used for further developments in the field by researchers and also, be integrated into text editors under a publicly available license.¹

1 Introduction

Spelling error detection and correction are respectively the two tasks of detecting a spelling error and correcting it by suggesting possible alternatives. These two tasks are generally carried out by a spell checking system which is built at early stages of development of language technology tools for a given language. Moreover, the analysis of the morphological structure of words is of importance as a subjacent component of natural language processing (NLP) pipelines where the morphemes of a given word are analyzed following morphological segmentation and tagging of composing morphemes. This task is beneficial for various purposes, particularly part-of-speech tagging, lemmatization and syntactic analysis.

Despite the recent advances in applying language-independent statistical and neural network methods, such as n-gram and sequence-to-sequence language modeling (Moeller et al., 2018; Ahmadi, 2018; Beemer et al., 2020), rule-based methods are still highly in use thanks to accuracy and viability, particularly in the case of morphologically rich languages like Kurdish. Among the rule-based methods, finite state transducers have been widely used for this task, as in (Fransen, 2020) for Old Irish, (Creutz and Lagus, 2002) for Finnish and (Nicholson et al., 2012) for Inuktitut. Although the tasks of morphological analysis, stemming and spell checking are addressed in previous studies for Kurdish, there is no open-source tool available (Ahmadi, 2020b). Therefore, in this paper, we present a rule-based approach to create a Sorani Kurdish morphological analyzer and spell checking system using HUNSPELL. HUNSPELL (Ooms, 2017) is a spell checker and morphological analyzer originally designed for languages with rich morphology and complex word compounding. An an open-source software, it is widely integrated by various web browsers and text editors for spell checking.

We previously described the Sorani Kurdish morphology from a computational point of view where the word-forms of lexemes with an open-class part-of-speech are defined (Ahmadi, 2021). These morphological rules are considered as finite-state transducers which can be used to analyze and generate various word-forms for a given lexeme. Using these rules, we define 4,293 morphological rules and demonstrate that the extracted rules can efficiently detect and suggest spelling errors and also, morphologically analyze words by stripping affixes, retrieving stems and extracting morphemes. Furthermore, we create a morphosyntactically-tagged lexicon containing 23,223 entries in Sorani Kurdish. The tagged lexicon along with the morphological rules are then used in HUNSPELL.

¹https://github.com/sinaahmadi/KurdishHunspell
The rest of the paper is organized as follows. We first briefly discuss the previous efforts in the same field in Section 2. In Section 3, the essential features of Sorani Kurdish morphology are described. In Section 4, we describe the development of the lexicon and the implementation of an affix file in Hunspell based on morphological rules to detect word forms. We present our experiments on spell-checking, stemming and morphological analysis in Section 5. Finally, the paper concludes in Section 6 where the main contributions of the current work along with a few suggestions for the future work are highlighted.

Throughout this manuscript, we use the Leipzig glossing rules (Comrie et al., 2008) for interlinear morpheme-by-morpheme glosses. We also mark the clitic and affix boundaries respectively with ‘=’ and ‘-’. Kurdish words are also provided in both Arabic-based and Latin-based scripts.

2 Related Work

2.1 Morphological Analysis

Some aspects of Kurdish morphology and syntax have been previously studied by linguists. Excluding pedagogical resources describing general grammar of Sorani Kurdish, we can mention the following major contributions to the field. To further clarify, an overview of Kurdish linguistics is provided by Haig and Matras (2002).

Among the analytic resources describing Sorani Kurdish morphology, we can mention (Hacî Marif, 2000; Blau, 2000; McCarus, 1958) where various morphological and syntactical features of Sorani Kurdish dialects are studied. On the interaction of morphology and phonology, Abdulla (1980) and Rashid Ahmad (1987) analyze some aspects of the Kurdish phonology where various morphological rules are taken into account. One of the specific characteristics of Kurdish is ergativity. This topic has been also widely studied in Indo-Iranian languages in general (Karimi, 2012), and Kurdish in particular (Jukil, 2015; Mahalingappa, 2013; Karimi, 2010; Karimi, 2014). Another interesting topic is the Ezafe (or Èafa) construction which has been addressed as well (Karimi, 2007; Sahin, 2018).

Focused on specific morphemes, Baban (2001) studies the repetition suffix -ewe in Sorani Kurdish and analyzes both morphology and syntax with respect to this morpheme. In the same vein, Doostan and Daneshpazhouh (2019) focus on the -râ morpheme from syntactic and semantic perspectives. Another related work to the current project is (McCarus, 2007) which analyzes the inflectional and derivational patterns in the Sorani Kurdish morphology. Fatah and Hamawand (2014) provide descriptions of some of the Sorani Kurdish positive and negative prefixes based on the Prototype Theory. Karacan and Khalid (2016) analyze adjectives in various Kurdish dialects. Walther (2012) investigates morphological structures of Sorani Kurdish mobile person markers. The work extends the formal analysis of Sorani Kurdish endoclitic placement by Samvelian (2007).

Although the knowledge of morphology for creating linguistic resources is of fundamental importance in NLP and computational linguistics, there are only a limited number of studies that deal with Kurdish computational morphology, directly or explicitly within other tasks. As the earliest attempt in generative grammar for Kurdish, Baban and Husein (1995) provide a description to create Kurdish sentences by focusing on both morphology and syntax. Despite its originality and novelty, this study does not seem to be widely used or followed by the succeeding studies. As such, Walther and Sagot (2010) and Walther et al. (2010) present a methodology and preliminary experiments on constructing a morphological lexicon in the Alexina Framework (Sagot, 2010) for both Sorani and Kurmanji in which a lemma and a morphosyn- tactic tag are associated with each known form of the word. The lexicon$^2$ contains 22,990 word forms of 274 lemmata and is added to the Universal Morphology (UniMorph) project$^3$ which is a collaborative effort to improve how NLP handles complex morphology in the world’s languages (Kirov et al., 2018; Cotterell et al., 2017). One limitation regarding this study is the lack of coverage of complex morphological forms such as inflected verbs based on tenses and aspects. Moreover, we noticed that many of the word forms are not correct and some follow the morphology of a specific sub-dialect of Sorani Kurdish and are not globally used in the dialect. More recently, Ahmadi and Hassani (2020) present a preliminary

---

$^2$https://github.com/unimorph/ckb
$^3$https://unimorph.github.io
study on creating a morphological analyzer using finite-state transducers. Furthermore, (Ahmadi, 2020a) creates a morphological analyzer for the task of lexical analysis, also known as tokenization, for both Sorani and Kurmanji dialects of Kurdish where the extraction of morphemes is taken into account as a sub-task.

### 2.2 Spelling Error Correction and Stemming

Similarly, this task has been previously addressed in a few studies. Salavati and Ahmadi (2018) present a system for lemmatization and spelling error correction using n-gram models and a non-exhaustive list of the most frequent morphemes in Sorani Kurdish. Similarly, Hawezi et al. (2019) propose an algorithm for Sorani Kurdish spell checking using a lexicon and a set of patterns to define word forms. A given entry is then corrected by calculating the Levenshtein distance. Saeed et al. (2018a) and Mustafa and Rashid (2018) propose a stemming system for Sorani Kurdish which retrieves word stems by removing longest suffix and prefixes of words. Salavati et al. (2013) present a rule-based and statistical stemmer for both Sorani and Kurmanji Kurdish. Saeed et al. (2018b) utilizes the Porter stemming algorithm for Kurdish text classification. The previous studies on stemming are carried out in the context of information retrieval and are reportedly efficient in stemming queries. More importantly, Amani and Koochari (2017) and Mahmudi (2019) create spell checkers for Sorani, respectively, using the Soundex algorithm and a statistical approach, and report their efficiency when evaluated on their test sets. The current study is compared with these two spell checking approaches.

Based on our literature review, no morphological analyzer or spelling error detection and correction system is openly available for Sorani Kurdish.

### 3 Sorani Kurdish Morphology

#### Bound morphemes:

| Clitics | Enclitics: =im, =î, =út, =e, =ê, =ët, =în, =î, =wa, =wat, =a, =at, =te |
|--------|--------------------------------------------------------------------------|
|        | Endoclitics: =im, =ît, =î, =man, =tan, =yan, =îş                       |
| Inflectional | Noun: -êk, -an, -gel, -ha, -at, -eke, -ekan, -e, -ane, -gele, -i, -ê, -îne |
| Adjective: -êr, -în |
| Adverb: -âne |
| Derivational: -a, -and, -ard, -d, -îšt, -ird, -t, -û, -y, -ên, -ër, -diran, -ran, -ewe, ber, da, der, hel, ser, êa, řo, wer, -û |

Figure 1: A classification of Sorani Kurdish bound morphemes with a few instances

Kurdish is an Indo-European language spoken by over 25 million speakers in the Kurdish regions in Turkey, Iran, Iraq and Syria, and also by the Kurdish diaspora around the world (McCarus, 2007). Sorani Kurdish, also known as Central Kurdish (ckb in ISO 639-2), is the Kurdish dialect which is mostly spoken by the Kurds within the Iranian and Iraqi regions of Kurdistan. Sorani Kurdish is a null-subject language and has a subject-object-verb (S-O-V) order and can be distinguished from other Indo-Iranian languages by its ergative-absolutive alignment which appears in past tenses of transitive verbs (Ahmadi and Masoud, 2020). In this section, we provide a brief description of the Sorani Kurdish morphology by focusing on morphemes. Morphemes are classified into free and bound. While free morphemes are meaningful as they are, bound morphemes only carry meaning when affixed with other words.

---

4 As of September 2021
3.1 Bound Morphemes

Bound morphemes are classified into two categories of affixes and clitics. Affixes and clitics are similar in the way that they cannot constitute a word and they lean on a prosodic host, i.e. a word for stress assignment. Clitics can appear with hosts of various syntactic categories while affixes only combine with syntactically-related stems (Haspelmath and Sims, 2013). The clitics and affixes in Sorani Kurdish have been widely studied previously and have been shown to be challenging considering the general theory of clitics (W. Smith, 2014; Gharib and Pye, 2018). This problem is particularly observed with respect to the direct and oblique person markers which can appear in different positions within a word-form depending on the functionality. In this section, the clitics and affixes in Sorani Kurdish are described. Figure 1 provides the most frequent clitics and affixes in Sorani Kurdish.

3.1.1 Clitics

Clitics are categorized based on their position with respect to the host. A clitic is called proclitic and enclitic, if it appears before and after the host, respectively. There are two other forms of clitics which are non-peripherical and exist only among a few number of natural languages. If a clitic appears between the host and another affix, it is called a mesoclitic. A different type of non-peripheral clitic is endoclitic which appears within the host itself and is unique to a few languages around the world, such as Udi (W. Smith, 2014), Degema (Kari, 2002) and also Sorani Kurdish.

Sorani Kurdish has two types of endoclitics: pronominal makers (also introduced as mobile person markers by (Walther, 2012)) and the emphasis endoclitic =îş which can be translated as ‘also’ or ‘too’. The pronominal endoclitics function as agent markers for transitive verbs in the past tenses or endoclitics as patient marker for transitive verbs in the present tenses. This is due to the split ergativity feature of Sorani Kurdish where the agent and patient markers are specified differently. The following examples show the alignment in present and past tenses of کوتن (kewtin, ‘to fall’) and گرن (girtin, ‘to get’). The agent marker -in in intransitive present tenses serve as patient marker in transitive past tenses due to ergativity.

(1) \textit{dekewin} دەکەوتن \linebreak[2] de-kew-in \linebreak[2] fall.prs.prog.intr.3pl \linebreak[2] ‘(they) are falling.’

(2) \textit{deghiri} دەگرتن \linebreak[2] de-gir-in \linebreak[2] get.prs.prog.tr.3pl \linebreak[2] ‘(they) are getting.’

(3) \textit{kewtin} کوتن \linebreak[2] kewt-in \linebreak[2] fall.pst.prog.intr.3pl \linebreak[2] ‘(they) fell.’

(4) \textit{girtyanin} گرتیانن \linebreak[2] girt=yan-in \linebreak[2] get.pst.prog.tr.erg.3pl.3pl \linebreak[2] ‘(they) got (them).’

Furthermore, the two endoclitic categories of Sorani appear in an erratic pattern within a word form or a phrase. Table 1 presents an example where the 1sg marker \textit{im} (=im) and the emphasis endoclitic =îş

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
|  |  | girt | im |  |  | ne |  | da |
|  |  | girt | im |  |  | îş | im | ne |
|  |  | girt | im | in |  |  | e | de |
|  |  | girt | im | in | e | ewe |  |  |
|  |  | girt | îş | im | in | e | ewe |  |  |
|  |  | girt | îş | im | in | e | ewe |  |  |
|  |  | girt | îş | im | in | e | ewe |  |  |
|  |  | girt | îş | im | in | e | ewe |  |  |

\begin{itemize}
\item \textit{girt} \textit{im} =îş \textit{im} \textit{e} \textit{ewe} \textit{girt} \textit{im} \textit{e} \textit{ewe} \textit{girt} \textit{îş} \textit{im} \textit{e} \textit{ewe} \textit{girt} \textit{îş} \textit{im} \textit{e} \textit{ewe} \textit{girt} \textit{îş} \textit{im} \textit{e} \textit{ewe} \textit{girt} \textit{îş} \textit{im} \textit{e} \textit{ewe} \textit{girt} \textit{îş} \textit{im} \textit{e} \textit{ewe}
\end{itemize}

Table 1: The placement of the endoclitic =îş (in green boxes) and agent marker =im (in blue boxes) with respect to the base and each other in a verb form. Note that Sorani Kurdish is a null-subject language.

Furthermore, the two endoclitic categories of Sorani appear in an erratic pattern within a word form or a phrase. Table 1 presents an example where the 1sg marker \textit{im} (=im) and the emphasis endoclitic =îş
appear after and before the host, i.e. گرت (girt), depending on the presence of other bound morphemes such as negation prefix ن (ne-) or the verbal particle ِا (da). It is worth mentioning that this pattern may vary based on the Sorani sub-dialects.

3.1.2 Affixes
In comparison to clitics, a bigger number of bound morphemes in Sorani Kurdish belong to affixes. Affixes can be categorized into inflectional and derivational based on their ability to create new lexemes. The most frequent affixes in Sorani appear as prefix and suffix. Some of the inflectional affixes of Sorani Kurdish belonging to open-class parts of speech, namely nouns, verbs, adjectives and adverbs, are shown in Table 2. In addition, Izafa particle -î and its allomorph -e which appear between a head and its dependents in a noun phrase are frequently used to create possessive constructions, as in نێوی من (nêw-î min 'my name').

| Nouns       | Verbs       | Adjectives | Adverbs    |
|-------------|-------------|------------|------------|
| number (SG, PL) | number (SG, PL) | number (SG, PL) | degree (COMP, SUPL) |
| person (1, 2, 3)  | person (1, 2, 3)   | person (1, 2, 3)  | degree (COMP, SUPL)  |
| determiners (DEF, IND, DEM) | mood (IND, SBJV, IMP, COND) | determiners (DEF, IND, DEM) | |
| case (OBL, LOC, VOC) | aspect (PRF, IMP, PROG) | | |
| gender (M, F) | tense (PST, PRS) | | |

Table 2: Inflectional features and values of Sorani Kurdish. It should be noted that the function of cases and genders vary among Sorani subdialects.

In addition to compound words, Sorani Kurdish relies on derivational morphemes to create new lexemes, particularly new verbal lexemes. To this end, verbal suffix ئوە (-ewe) and verbal particles such as ِا (da-) and ِه (heł-) are used. It is worth noting that passive form of verbs is derived from the verb stem by using دران/درێ (-diran/-dirê) (or their allomorphs دران/دێرئ (-ran/-rê) (Doostan and Daneshpazhouh, 2019)) suffixes, unlike Kurmanji Kurdish which relies on periphrastic forms with هاتین (to come).

A more detailed description of Sorani Kurdish morphology, including adpositions and pronouns as free morphemes is provided in (Ahmadi, 2021).

4 Implementation
As a rule-based method, Hunspell needs an annotated lexicon to which the morphological rules are applied. The lexicon (with .dic extension) contains words with morphosyntactic flags, and the morphological rules are included in the affix file (with .aff extension) that specifies how word forms are created based on the word flags in the lexicon. The creation of the lexicon and implementation of the rules are described in this section and a few samples of these two files are provided in Appendix A.

4.1 Lexicon Annotation
Since there is no annotated lexicon available for this purpose, we use the Sorani Kurdish lexicographic material provided by Wikiferheng, the Kurdish Wiktionary5 and the FreeDict project6. The latter provides the content in the Latin-based script of Kurdish. Therefore, we use the rule-based transliteration system provided by (Ahmadi, 2019) to transliterate it into the Arabic-based script which is used in our implementation. In addition, Wikidata is consulted to extract proper names using the query provided in Appendix B. Each lemma in the lexicon is manually tagged with information such as part-of-speech, its formation type, i.e. derivational or inflectional, and further morphological properties such as stem, person and tense. In addition, composing parts of compound forms are specified using a hyphen which can be beneficial for lexical analysis. The number of entries per morphosyntactic tags are provided in Table 3.

5https://ku.wiktionary.org
6https://freedict.org
4.2 Morphological Rules

The morphological rules in Hunspell are defined as prefixes and suffixes, respectively specified with PFX and SFX, in such a way that new forms of the lemmata and stems in the lexicon are analyzed or generated by stripping these morphemes from a given word-form or adding them to the lemmata or stems in the lexicon. In addition, it is possible to integrate the morphophonological features which create the surface realization of word forms based on the phonology of the language. Following the morphophonological rules described in (Ahmadi, 2021), automatic and morphophonological alternations are integrated in the morphological rules in the affix file. Although Hunspell allows specifying automatic alternations in brackets ([ ]) by defining different rules for a specific suffix if preceded by a vowel or consonant, as provided in Appendix A, morphophonological alternations are defined as separate rules and in some cases may require separate entries in the lexicon as well. For instance, the morpheme رەوە (rewe) in تەوێزەرەوە (bixwênerewe, read.v. imp.2sg) is a morphophonological alternation of the morpheme ەوە (ewe) which is therefore defined as a new rule.

Moreover, although Hunspell allows twofold prefix stripping, we only define single stripping rules for both prefixes and suffixes to facilitate the maintenance and future developments. In other words, the set of possible combinations of prefixes or suffixes that can appear with a base are first detected and then defined in Hunspell as individual rules. Table 4 summarizes the number of rules in our implementation based on their morphosyntactic categories. It should be noted that prefix and suffix in this context refer to the position of the morphemes with respect to the host. Therefore, clitics are also included in the rules similar to affixes.

In order to be able to use Hunspell for morphological analyzer, we use a wrapper program\(^7\).

| Prefix | Suffix |
|--------|--------|
| N      | 20,538 |
| V      | 1,005  |
| I      | 779    |
| T      | 271    |
| A      | 1,732  |
| R      | 441    |
| E      | 51     |
| C      | 38     |
| B      | 82     |
| F      | 30     |
| E      | 51     |
| F      | 13     |
| G      | 16     |
| X      | 3,528  |
| Z      | 9,893  |
| W      | 39     |
| All    | 23,223 |

Table 4: Statistics of the morphological rules for Sorani Kurdish

5 Experiments

5.1 Spelling Error Correction

Test Sets Sorani Kurdish lacks a unified orthography. To tackle this problem, we follow the guidelines provided by (Hashemi, 2016) as our orthographic reference. These guidelines are similarly taken into account in the two test sets which are used by (Amani and Koochari, 2017) and (Mahmudi, 2019). Therefore, we use these test sets and respectively refer to them as T\(_1\) which contains 385 instances among which there is no correctly spelled word, and T\(_2\) which contains 1,400 instances among which %60.6 are

\(^7\)https://github.com/MSeal/cython_hunspell
correct. Figure 2 illustrates a description of this dataset. In addition, T₂ provides more challenging incorrectly spelled instances, such as those due to missed spaces, i.e. two or more separate words merged together without a space. Such cases are specified as “incorrect (spaced)” in Figure 2. We note that the spelling errors in T₂ contain more inflected forms such as plural forms, conjugated verbs and various verb tenses in comparison to T₁.

Baseline System  Furthermore, we extract the 2,376,405 unique word forms from Sorani Kurdish corpora, namely those provided by (Esmaili et al., 2013; Veisi et al., 2020; Ahmadi et al., 2020) among which words with a frequency of 10 or more are considered as potentially correct words that can be used to create a baseline system. Such words make almost 11% of the whole extracted words, i.e. 265,216 words. In the baseline system, spelling errors are detected by looking up in the list of the words and also corrected by calculating the Levenshtein distance (Levenshtein, 1966) of the word with the unique words from the corpora. This way, if an input word appears in the list of the frequent words, it is considered correct, otherwise the baseline system provides suggestions sorted based on the edit distance.

Evaluation  Table 5 provides the performance of the baseline system and our implementation in Hunspell along with the reported performance of the previous systems.⁸ We calculate accuracy (Acc), Lexical Precision (Pₜ), Lexical Recall (Rₜ) and F₁ measures as defined by (Van Huyssteen et al., 2004) and based on the followings:

- true positive (TP) for correctly detected correct words.
- false positive (FP) for incorrectly detected correct words.
- true negative (TN) for correctly detected incorrect words.
- false negative (FN) for incorrectly detected incorrect words.

In addition, we calculate the existence of a correct prediction of an incorrect word based on their ranking, as in the first, three and more than three first suggestions of the systems. These cases are respectively denoted by Sugg₁, Sugg₃ and Sugg₃. It should be noted that the correction suggestions are ranked by the most probable suggestions, i.e. the most probable replacement is supposed to be suggested first.

Our implementation in Hunspell performs differently and disparately with respect to each test set with an overall efficiency below the baseline system and that of the previous systems. Quantitatively, in all cases, our system fails to significantly detect correct words correctly leading to a lower precision, and also, fails to remarkably detect correct words among all the words, leading to a lower recall. On the other hand, the accuracy also indicates the competency of the spell checker in handling a given word accurately, i.e. the likelihood that the suggested words of the spell checker for a correct or incorrect input

---

⁸Please note that the evaluation of our system is carried out on Version 0.1.0. See https://github.com/sinaahmadi/KurdishHunspell
are correct. In the current version of our implementation, Hunspell is only accurate in 40% of the cases. Nevertheless, it is not clear to us whether the reported lexical precision and recall in the previous studies are calculated based on the prediction of correct instances or incorrect ones. Figure 3 shows a comparison of our system along with the baseline and the previous systems based on F1 score and the existence of a correct prediction in the suggestions, i.e. Sugg\_>3.

| System       | Test set       | TP   | FP   | TN   | FN   | Acc (%) | P_r | R_c | F1   | Sugg_1 (%) | Sugg_3 (%) | Sugg_>3 (%) |
|--------------|----------------|------|------|------|------|---------|-----|-----|------|------------|------------|-------------|
| Baseline     | T1             | 0    | 0    | 178  | 206  | 26.82   | 0.70| 0.58 | 0.63 | 10.87      | 15.76      | 19.02       |
|              | T2             | 597  | 251  | 105  | 441  | 50.36   | 0.70| 0.74 | 0.72 | 18.69      | 27.10      | 32.71       |
| (Amani and   | T1             | 0    | 0    | 208  | 176  | 24.30   | 0.98| 0.97 | 0.98 | 68.17      | 73.94      | 74.94       |
| Koochari,    | T2             | 357  | 108  | 753  | 0    | 92.29   | 0.83| 1.00 | 0.91 | 69.57      | 80.33      | 84.97       |
| 2017)\(^9\)  | T2\_space     | 367  | 481  | 182  | 364  | 39.38   | 0.43| 0.50 | 0.46 | 18.30      | 26.99      | 31.52       |
| (Mahmudi,    | T1             | 0    | 0    | 208  | 176  | 22.92   |     |     |     |            |            |             |
| 2019)\(^9\)  | T2             | 359  | 108  | 753  | 0    | 92.29   | 0.83| 1.00 | 0.91 | 69.57      | 80.33      | 84.97       |
| Our system   | T1             | 0    | 0    | 208  | 176  | 22.92   |     |     |     |            |            |             |
|              | T2             | 367  | 481  | 182  | 364  | 39.38   | 0.43| 0.50 | 0.46 | 18.30      | 26.99      | 31.52       |

Table 5: The performance of our implementation in Hunspell in comparison with the reported performance of the previous studies. The gray cells correspond to unreported scores in the related study or undefinable ones.

Figure 3: A comparison of the performance of our systems along with the two previous studies based on the F1-score and the existence of a correct prediction, i.e. Sugg\_>3

From a qualitative point of view, the results of the evaluation of our system can be explained by the following issues:

(a) non-existence of a word of the test set in the annotated lexicon, e.g. 'هۆڵیوود' 'Hollywood'. It should be noted that there are non-Kurdish words in the test sets, for instance words in Gorani or Arabic, which does not exist in our Sorani lexicon and therefore, cannot be correctly predicted.

(b) incorrect morphological synthesis of the input word or incorrect morphological generation of a suggested word, e.g. 'دهواین' (det wayn) as an incorrect suggestion for 'دەتواین' (detwain) instead of 'دەتواین' (detwain, 'we can').

(c) lack of lexical variations in the test sets, e.g. 'تەنیا' (tenya, 'alone (adjective)') is considered an incorrect spelling of 'تەنیا' (tena) in (Mahmudi, 2019)’s test set, while both words are used in various sub-dialects of Sorani and thus, exist in our lexicon and are detected as correct.

(d) different convention regarding spacing as in 'دەریبکەن' (deribiken, 'that you.plural) dismiss’) which is a compound verb and can be also written with spaces in between composing parts as 'دەریبیکن' (deri biken). However, such spacing is deemed incorrect in (Mahmudi, 2019)’s test set, while our system detects it as a correct spelling. Mahmudi (2019)’s test set is curated with the assumption that all compound forms in Kurdish are written contiguous without any spaces.

\(^9\)The results of this system is based on the reported scores in (Mahmudi, 2019)

\(^{10}\)See footnote 9
To further examine the last aforementioned point, we evaluate our baseline and Hunspell spell checkers on those words in T\_2 for which spelling errors are not due to missing spaces, i.e. the red and green portions of Figure 2. This test set is referred to as T\_2\_space in Table 5. Unlike the baseline system, the results indicate that Hunspell performs with the same precision of 0.43 but the higher recall of 0.62 when tested on T\_2\_space in comparison to T\_2. This can be explained by the lower number of incorrect words in T\_2\_space which increases the likelihood of detecting an input word as correct (73.50% correct words in T\_2\_space in comparison to 61.4% in T\_2). Given the insignificance of difference between other scores, Sugg\_2\_3 seems to be the only representative measure that indicates a lower percentage of correct suggestions (27.73%) by Hunspell on T\_2\_space that is lower than Sugg\_2\_3 of T\_2 (31.52%).

Based on this qualitative analysis, we believe that our systems carry out the task of spell checking efficiently but differently from the expected corrections in the test sets, particularly in the case of compound words. It is worth mentioning that the Arabic-based scripts are known to present many challenges in spelling complexity (Habash et al., 2018). Ahmadi (2020a) tackles inconsistencies of spacing for Kurdish compound words differently for lexical analysis. Moreover, one major limitation of Hunspell is ignoring spaces in the input words. For instance, if the input word بەهار (behar ‘spring (noun)’) is mistakenly split into two tokens بە (be) and هار (har), the spell checker cannot detect that the space causes an error. Given that there is no such type of error in the test sets, we believe that this feature of our implementation in Hunspell should be evaluated with a more representative test set.

Appendix C provides a few examples of the outputs of our baseline system and the implementation in Hunspell. To better describe the type of errors that the systems made, various types of incorrect words and predictions are specified in different colors.

5.2 Morphological Analysis

Although the Unimorph project provides a dataset for Sorani Kurdish morphological analysis\(^{11}\), this dataset contains many incorrect word forms which seem to be generated without sufficient knowledge to assess Sorani Kurdish morphotactics. Therefore, we create a new test set as a preliminary effort to analyze Sorani Kurdish morphology. We annotate 140 word forms with information such as lemma, part-of-speech tag, stem, base, prefixes, suffixes and additional description. In this context, prefixes and suffixes respectively refer to the preceding and succeeding segments that appear before and after the base. Among the 140 word forms, 14 are adjectives, 16 are adverbs, 27 are nouns, 73 are verbs and the remaining 10 words have two or more part-of-speech tags.

We evaluate our morphological analyzer in Hunspell from the following three perspectives:

- Morphological segmentation which breaks a word into its composing morphemes (Poon et al., 2009). Evaluation results indicate that Hunspell has an accuracy of 80.14% to retrieve the base of a given word along with its preceding and succeeding morphemes.
- Analysis coverage: based on the unique word forms extracted from the corpora (discussed in Section 5.1), we calculate the number of word forms that can be segmented by our morphological analyzer, correctly or incorrectly. This indicates the extent to which morphological structures are known to the morphological analyzer and therefore, it can reveal the ability of the system to deal with the complexity of word forms. To this end, the extracted forms from the corpora are tokenized and preprocessed using KLPT (Ahmadi, 2020b). Among the 235,210 unique word forms, the morphological analyzer can analyze 27.16% of the words. We believe that this is due to a lack of lexical entries in our lexicon, and also many morphological patterns particularly from specific sub-dialects of Sorani Kurdish. Nevertheless, not all the extracted word forms are necessarily correctly spelled or valid.
- Part-of-speech tagging and stemming: Our implementation in Hunspell assigns a correct part-of-speech tag with an accuracy of 86.02%. Likewise, the stemming of verbs is carried out with an accuracy of 63.75%.\(^{12}\)

---

\(^{11}\) Available at https://github.com/unimorph/ckb

\(^{12}\) Only stemming of verbs is possible in the current version.
We believe that the lower performance of the morphological analysis is due to undefined morphological rules of more challenging word forms such as نەشیانبینیبوم (nesyanbinibum, ‘(they) even had not seen me’), where the endoclitic =y appears before the negation prefix ne- and the personal marker =yan (bîn is the host in this word). In addition, some word forms are more ambiguous to be analyzed. For instance, چیکوشانم (têkoşanim, ‘my efforts’) is incorrectly tagged as a verb by Hunspell due to the host being the infinitive چیکوشان (têkoşan, ‘to try’). Likewise, compound words are error-prone due to the space, as in پێیدا (pêyda) which is an absolute adposition but is analyzed similar to پـێـیدا (pêy da, ‘he/she gave it to him/her’). It is worth mentioning that the lemmas of the word forms in our test set exist in the lexicon, making the inaccuracy due to unseen words less significant.

6 Conclusion and Future Work

In this paper, a morphological analyzer and spell checker are introduced for Sorani Kurdish using Hunspell. Although morphological analysis and spell checking have been previously addressed for Sorani Kurdish, the literature review indicates that no tool is publicly available. As such, the current study describes tremendous efforts for defining morphological rules and tagging a lexicon to create an efficient and open-source tool for Sorani Kurdish. We demonstrate that our implementation performs efficiently for morphological analysis, spelling error detection and correction and also stemming. Given the complexity of Sorani Kurdish morphology, we believe that our tool and resources are useful to pave the way for further developments in Kurdish language processing. We also hope that our tool raises awareness among Kurdish users regarding writing and ultimately, help the tool to facilitate the standardization of the Sorani Kurdish orthography.

One major limitation in the tasks of morphological analysis and spelling error correction of Kurdish is the lack of benchmarking. This should be addressed by creating evaluation datasets that distinguishes between sub-dialects of Sorani, particularly Northern and Southern Sorani, and provides a finer representation of different spelling errors, especially for words in context rather than tokens. Therefore, as future work, the creation of a gold-standard dataset for evaluating spell checking systems for Kurdish is required. Moreover, addressing the morphology of subdialects of Sorani Kurdish and also, other variants of Kurdish to create similar tools and resources is necessary. We believe that our test set for evaluating Sorani Kurdish morphology should be further enriched with more complex word forms in a comprehensive way. We suggest that more derivational and inflectional morphemes to be added to the current implementation and also, the lexicon be further enriched over time. Based on the current study, more advanced NLP tasks, such as syntactic analysis, can be addressed in the future.

It is worth noting that both the spell checker and the morphological analyzer are openly available and operational in the Kurdish Language Processing Toolkit.13

7 Acknowledgements

The author would like to greatly thank those who supported open-source projects in Kurdish language technology through GitHub Sponsors.14

---

13https://github.com/sinaahmadi/klpt
14https://github.com/sponsors/sinaahmadi
References

[Abdulla1980] Jamal Jalal Abdulla. 1980. Some aspects of language purism among Kurdish speakers. Ph.D. thesis, University of York.

[Ahmadi and Hassani2020] Sina Ahmadi and Hossein Hassani. 2020. Towards Finite-State Morphology of Kurdish. arXiv preprint arXiv:2005.10652.

[Ahmadi and Masoud2020] Sina Ahmadi and Maraim Masoud. 2020. Towards Machine Translation for the Kurdish Language. In Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages, pages 87–98.

[Ahmadi et al.2020] Sina Ahmadi, Hossein Hassani, and Kamaladdin Abedi. 2020. A corpus of the Sorani Kurdish folkloric lyrics. In Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL), pages 330–335.

[Ahmadi2018] Sina Ahmadi. 2018. Attention-based encoder-decoder networks for spelling and grammatical error correction. arXiv preprint arXiv:1810.00660.

[Ahmadi2019] Sina Ahmadi. 2019. A rule-based Kurdish text transliteration system. Asian and Low-Resource Language Information Processing (TALLIP), 18(2):18:1–18:8.

[Ahmadi2020a] Sina Ahmadi. 2020a. A Tokenization System for the Kurdish Language. In Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects, pages 114–127.

[Ahmadi2020b] Sina Ahmadi. 2020b. KLPT–Kurdish Language Processing Toolkit. In Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS), pages 72–84.

[Ahmadi2021] Sina Ahmadi. 2021. A Formal Description of Sorani Kurdish Morphology. arXiv preprint arXiv:2109.03942.

[Amani and Koochari2017] Arash Amani and Abbas Koochari. 2017. Improving Kurdish spelling error correction by modifying minimum edit distance measures. In 2nd National conference on Soft Computing, Gilan, Iran.

[Baban and Husein1995] ST Baban and S Husein. 1995. Programmable Grammar of the Kurdish Language. ILLC Research Report and Technical Notes Series, University of Amsterdam.

[Baban2001] Sherko Baban. 2001. The grammar of the “ewe” suffix in Sorani Kurdish (Rêzmanî paşgirî diâpati “ewe”) [in Kurdish]. Mêkîrîyanî Establishment For Research & Publication.

[Beemer et al.2020] Sarah Beemer, Zak Boston, April Bukoski, Daniel Chen, Princess Dickens, Andrew Gerlach, Torin Hopkins, Parth Anand Jawale, Chris Koski, Akanksha Malhotra, et al. 2020. Linguist vs. machine: Rapid development of finite-state morphological grammars. In Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, pages 162–170.

[Blau2000] Joyce Blau. 2000. Méthode de kurde: sorani. Editions L’Harmattan.

[Comrie et al.2008] Bernard Comrie, Martin Haspelmath, and Balthasar Bickel. 2008. The Leipzig glossing rules: Conventions for interlinear morpheme-by-morpheme glosses. Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology & the Department of Linguistics of the University of Leipzig. Retrieved January, 28:2010.

[Cotterell et al.2017] Ryan Cotterell, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sandra Kübler, David Yarowsky, Jason Eisner, et al. 2017. Conll-sigmorphon 2017 shared task: Universal morphological reinflection in 52 languages. arXiv preprint arXiv:1706.09031.

[Creutz and Lagus2002] Mathias Creutz and Krista Lagus. 2002. Unsupervised discovery of morphemes. arXiv preprint cs/0205057.

[Doostan and Daneshpazhouh2019] Gholamhossein Karimi Doostan and Fatemeh Daneshpazhouh. 2019. Kurdish -rd as an Anti-Actormarker, volume 1. University of Bamberg Press.

[Esmaili et al.2013] Kyumars Sheykh Esmaili, Donya Eliassi, Shahin Salavati, Purya Aliabadi, Asrin Mohammadi, Somayeh Yosefi, and Showmen Hakimi. 2013. Building a test collection for Sorani Kurdish. In Computer Systems and Applications (AICCSA), 2013 ACS International Conference on, pages 1–7. IEEE.
[Fatahand Hamawand2014] Azad Hasan Fatah and Zeki Hamawand. 2014. A Prototype Approach to Kurdish Prefixes. *International Journal on Studies in English Language and Literature (IJSELL)*, 2(10):37–49.

[Fransen2020] Theodorus Fransen. 2020. Automatic morphological analysis and interlinking of historical Irish cognate verb forms. In *Morphosyntactic Variation in Medieval Celtic Languages*, pages 49–84. De Gruyter Mouton.

[Gharib and Pye2018] Hiba Gharib and Clifton Pye. 2018. The clitic status of person markers in Sorani Kurdish. *University of Kansas Department of Linguistics*.

[Habash et al.2018] Nizar Habash, Fadhil Eryani, Salam Khalifa, Owen Rambow, Dana Abdulrahim, Alexander Erdmann, Reem Faraj, Wajdi Zaghouni, Houda Bouamor, Nasser Zalmout, et al. 2018. Unified guidelines and resources for Arabic dialect orthography. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

[Hacî Marif2000] Ewrehman Hacî Marif. 2000. Kurdish language grammar (Rêzmanî Zimanî Kurdî) [in Kurdish]. *Sardam Publishing House*, 1-4.

[Haig and Matras2002] Geoffrey Haig and Yaron Matras. 2002. Kurdish linguistics: a brief overview. *STUF-Language Typology and Universals*, 55(1):3–14.

[Hashemi2016] Dyako Hashemi. 2016. Kurdish orthography [In Kurdish]. [http://yageyziman.com/Renusi_Kurdi.htm](http://yageyziman.com/Renusi_Kurdi.htm). Accessed: 2020-07-25.

[Haspelmath and Sims2013] Martin Haspelmath and Andrea D Sims. 2013. *Understanding morphology*. Routledge.

[Hawezi et al.2019] Roojwan Sc Hawezi, Muhammed Y Azeez, and Ahmed A Qadir. 2019. Spell checking algorithm for agglutinative languages “Central Kurdish as an example”. In *2019 International Engineering Conference (IEC)*, pages 142–146. IEEE.

[Jukil2015] Ali M Jukil. 2015. Ergativity in Kurdish language. *Journal of Humanity Sciences*, 19(6):1–22.

[Karacan and Khalid2016] Hasan Karacan and Hewa Salam Khalid. 2016. Adjectives in Kurdish language: Comparison between dialects. *International Journal of Kurdish Studies*, 2(2).

[Kari2002] Ethelbert E Kari. 2002. On endoclitics: Some facts from Degema. *Journal of Asian and African Studies*, 63:37–53.

[Karimi2007] Yadgar Karimi. 2007. Kurdish Ezafe construction: Implications for DP structure. *Lingua*, 117(12):2159–2177.

[Karimi2010] Yadgar Karimi. 2010. Unaccusative transitives and the Person-Case Constraint effects in Kurdish. *Lingua*, 120(3):693–716.

[Karimi2012] Yadgar Karimi. 2012. The evolution of ergativity in Iranian languages. *Acta Linguistica Asiatica*, 2(1):23–44.

[Karimi2014] Yadgar Karimi. 2014. On the syntax of ergativity in Kurdish. *Poznan Studies in Contemporary Linguistics*, 50(3):231–271.

[Kirov et al.2018] Christo Kirov, Ryan Cotterell, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sebastian Mielke, Arya D McCarthy, Sandra Kübler, et al. 2018. Unimorph 2.0: Universal morphology. *arXiv preprint arXiv:1810.11101*.

[Levenshtein1966] V. I. Levenshtein. 1966. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10:707, February.

[Mahalingappa2013] Laura Mahalingappa. 2013. The acquisition of split-ergative case marking in kurmanji kurdish. *The Acquisition of Er-gativity (239-270). Amsterdam/Philadelphia: John Benjamins*.

[Mahmudi2019] Aso Mahmudi. 2019. Design and implementation of a spell checker for Kurdish (Central dialect). Master’s thesis at the University of Tehran, Iran.

[McCarus1958] Ernest N McCarus. 1958. *A Kurdish Grammar: Descriptive Analysis of the Kurdish of Sulaimaniya, Iraq*. American Council of Learned Societies Program in Oriental Languages.

[McCarus2007] Ernst M McCarus. 2007. Kurdish morphology. *Morphologies of Asia and Africa*, 2:1021–1049.
[Moeller et al. 2018] Sarah Moeller, Ghazaleh Kazeminejad, Andrew Cowell, and Mans Hulden. 2018. A neural morphological analyzer for Arapaho verbs learned from a finite state transducer. In Proceedings of the Workshop on Computational Modeling of Polyisynthetic Languages, pages 12–20.

[Mustafa and Rashid 2018] Arazo M Mustafa and Tarik A Rashid. 2018. Kurdish stemmer pre-processing steps for improving information retrieval. Journal of Information Science, 44(1):15–27.

[Nicholson et al. 2012] Jeremy Nicholson, Trevor Cohn, and Timothy Baldwin. 2012. Evaluating a morphological analyser of Inuktitut. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 372–376.

[Ooms 2017] Jeroen Ooms. 2017. hunspell: High-Performance Stemmer, Tokenizer, and Spell Checker. https://hunspell.github.io/.

[Poon et al. 2009] Hoifung Poon, Colin Cherry, and Kristina Toutanova. 2009. Unsupervised morphological segmentation with log-linear models. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 209–217.

[Rashid Ahmad 1987] Abdul-Majeed Rashid Ahmad. 1987. The Phonemic System of Modern Standard Kurdish. In PhD dissertation of the University of Michigan.

[Saeed et al. 2018a] Ari M Saeed, Tarik A Rashid, Arazo M Mustafa, Rawan A Al-Rashid Agha, Ahmed S Shamaldin, and Nawzad K Al-Salhi. 2018a. An evaluation of reber stemmer with longest match stemmer technique in kurdish sorani text classification. Journal of Computer Science, 1(2):99–107.

[Saeed et al. 2018b] Ari M Saeed, Tarik A Rashid, Arazo M Mustafa, Polla Fattah, and Birzo Ismael. 2018b. Improving Kurdish web mining through tree data structure and porter’s stemmer algorithms. UKH Journal of Science and Engineering, 2(1):48–54.

[Sagot 2010] Benoît Sagot. 2010. The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. In 7th international conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta, May.

[Sahin 2018] Robin Sahin. 2018. Extended Ezafe Particles in Tor Kurdish. Master’s Thesis in Iranian Languages, Department of Linguistics and Philology of University of Uppsala.

[Salavati and Ahmadi 2018] Shahin Salavati and Sina Ahmadi. 2018. Building a lemmatizer and a spell-checker for Sorani Kurdish. arXiv preprint arXiv:1809.10763.

[Salavati et al. 2013] Shahin Salavati, Kyumars Sheykh Esmaili, and Fardin Akhlaghian. 2013. Stemming for Kurdish information retrieval. In Asia Information Retrieval Symposium, pages 272–283. Springer.

[Samvelian 2007] Pollet Samvelian. 2007. What Sorani Kurdish absolute prepositions tell us about cliticization. Texas Linguistic Society IX: The morphosyntax of underrepresented languages, pages 263–283.

[Van Huysssteen et al. 2004] Gerhard B Van Huysssteen, E Roald Eiselen, and Martin J Puttkammer. 2004. Re-evaluating evaluation metrics for spelling checker evaluations. In Proceedings of First Workshop on International Proofing Tools and Language Technologies, pages 91–99.

[Veisi et al. 2020] Hadi Veisi, Mohammad MohammadAmini, and Hawre Hosseini. 2020. Toward Kurdish language processing: Experiments in collecting and processing the AsoSoft text corpus. Digital Scholarship in the Humanities, 35(1):176–193.

[W. Smith 2014] Peter W. Smith. 2014. Non-peripheral cliticization and second position in Udi and Sorani Kurdish. In Paper under revision at Natural Language and Linguistic Theory. (Date accessed: 12.05.2020), https://user.uni-frankfurt.de/~psmith/docs/smith_non_peripheral_cliticization.pdf edition.

[Walther and Sagot 2010] Géraldine Walther and Benoît Sagot. 2010. Developing a large-scale lexicon for a less-resourced language: General methodology and preliminary experiments on Sorani Kurdish. In Proceedings of the 7th SaLTML Worksho on Creation and use of basic lexical resources for less-resourced languages (LREC 2010 Workshop).

[Walther et al. 2010] Géraldine Walther, Benoît Sagot, and Karên Fort. 2010. Fast development of basic nlp tools: Towards a lexicon and a POS Tagger for Kurmanji Kurdish. In International Conference on Lexis and Grammar.

[Walther 2012] Géraldine Walther. 2012. Fitting into morphological structure: accounting for Sorani Kurdish endoclitics. In Mediterranean Morphology Meetings, volume 8, pages 299–321. [Online; accessed 19-Mar-2019].
Appendices

A Implementation

| po:verb is:present_stem_transitive_active st:ئاخ | SFX N 0 |
| po:verb is:present_stem_intransitive_passive st:ئاخێندر | SFX N 0 |
| po:verb is:present_stem_intransitive_passive st:ئاخێندران | SFX N 0 |
| po:verb is:infinite_intransitive_passive st:ئاخێو | SFX N 0 |

Table A1: Samples of lexicon entries to the right and morphological rules to the left in Hunspell for Sorani Kurdish. The po, is and st flags refer to part-of-speech tag, inflectional suffixes and stem, respectively. Morphophonological rules are provided in (Ahmadi, 2021).

B SPARQL Queries

The following queries are used to extract Sorani Kurdish labels of cities (Q515), countries (Q6256) and humans (Q5) on Wikipedia using the Wikidata SPARQL endpoint. ID in the query is to be replaced by concept identifiers.

```sparql
SELECT ?itemLabel
{
  ?item wdt:P31 wd:ID . # Q6256: country, Q5: human, Q515: city
  ?article schema:about ?item .
  ?article schema:isPartOf <https://ckb.wikipedia.org/> .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "ckb" . }
  FILTER( EXISTS { ?item rdfs:label ?lang_label. FILTER(LANG(?lang_label) = "ckb") })
} LIMIT 100
```

C Output of the Systems

The following table provides a few examples in the test sets ((Amani and Koochari, 2017; Mahmudi, 2019)) which are used for evaluating our baseline spell checker and our implementation in Hunspell. Various types of correct and incorrect words and suggestions are specified in different colors.
| Test set | Word | Correction | Our baseline | Suggestions |
|----------|------|------------|--------------|-------------|
|          | بزێن | detected as correct | detected as correct | Hunspell |
| (Amani and Koochari, 2017) | بوون | مانیا - بزێن - بوون - بوون - بوون - بوون - بزێن - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - بوون - Boo |