



Computational Modeling of Kashmiri Morphology: Finite-State Analysis and Linguistic Insights

¹Benish Afzal Want
Research Scholar, University of Kashmir
Email: ishbeenafzal@gmail.com

²Mudasir Mushtaq Pir
University of Kashmir

³Saduf Afzal
Kashmir University

Date of Submission: 15-03-2024

Date of Acceptance: 30-03-2024

Abstract

One of the important phases of Natural Language Processing is Morphological Analysis, enabling computers to understand and process human languages effectively and helps in machine translation. It is a tool used in NLP that dissects words into their smallest meaningful units or morphemes. Effective implementation of the morphological analyzer can be seen in language which is rich in morphemes. This paper focuses on the development of a morphological analyzer tailored for the Kashmiri language. We explore the linguistic characteristics of Kashmiri, discuss the architecture and design considerations of the morphological analyzer, address challenges in its development, and examine its potential applications in NLP tasks. By providing insights into the development of morphological analyzers for Kashmiri, this paper aims to contribute to the

advancement of NLP research and technology for under-resourced languages. This paper describes a two-level morphological analyzer for the Kashmiri language.

Key words: Natural language processing, Morphological analysis, Machine translation, Kashmiri language

I. Introduction:

Natural language understanding and processing have become integral parts of many applications, ranging from search engines and virtual assistants to machine translation and sentiment analysis. It is among the fields of study that is expanding the fastest. The four steps of natural language processing are as follows. Morphological analysis comes first, followed by syntactic, semantic, and lastly pragmatic analysis.

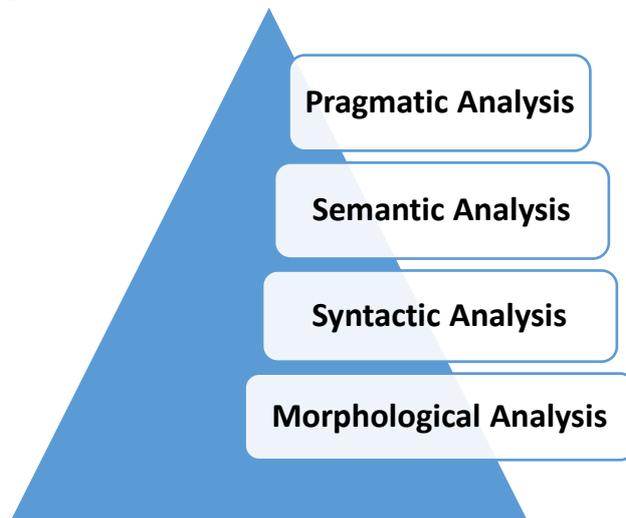


Fig 1: Steps in Natural Language Processing



Morphology delves into the internal structure of words and their meaningful components which is pivotal to understand languages. Morphological analysis entails breaking down words into the smallest meaningful units called morphemes. Each morpheme serves as the smallest building block in linguistic analysis, bearing both grammatical function and significance. Morphological analysis revolves around identifying the root, stem, and accompanying categorial details of linguistic elements. Various methodologies exist for this analysis within natural language processing, including the brute force method, root-driven approach, and affix stripping technique. These methods have notably evolved, particularly in analyzing Indian languages, to enhance precision in their analysis. Brute force stemmers utilize lookup tables, which map relations between root forms and inflected forms.

The two types of morphology in linguistic analysis are

- a) Inflectional Morphology
- b) Derivational Morphology

Inflectional morphology studies the systematic alterations made in words to convey grammatical meaning in a language. These modifications include modifications in the structure of a word to express variations in tense, aspect, mood, number, person, gender, case, and other grammatical categories. Inflectional morphology studies the rules which govern the formation and usage of these inflectional forms across different languages. It studies how languages employ inflectional affixes, infixes, suffixes, prefixes, and internal changes to express grammatical relationships and distinctions within sentences and discourse.

Consider the following example from the Kashmiri language:

/hu:n/ (dog) becomes /hu:ni/ (dogs) by adding the suffix 'i'.

/k^hjun/ (to eat) becomes /k^hjo:v/ (ate) and /k^hem/ (will eat)

/gindun/ (to play) becomes /gju:nd/ (played) and /gindi/ (will play)

Derivational morphology studies the formation and analysis of words through the addition of affixes or other morphological processes to existing base words. Unlike inflectional morphology, which mainly entails changes to signify grammatical relationships within a word, derivational

morphology is centered on the formation of new words or adjustments to the meaning or lexical category of established words. In derivational morphology, prefixes, suffixes, infixes, and circumfixes are appended to base words to generate derived forms imbued with novel meanings, functionalities, or lexical classifications. These affixes have the potential to alter the semantic essence, syntactic characteristics, or derivational categorization of the base word.

Example:

/ranun/ (to cook) becomes /rannawun/ (to make somebody cook) by adding the suffix '-wun'

/mo:l/ (father) becomes /mo:dʒ/ (mother) by final consonant change.

A morphological analyzer necessitates the analysis of each word in the lexicon and the establishment of morphological rules. In languages such as English and French, it has been noted that these rules can be entirely articulated through finite-state devices. These mechanisms are commonly employed in tackling morphological issues and consequently have developed into a distinct area of inquiry known as finite-state morphology. Likewise, the phrase "finite-state morphological analyzer" denotes a morphological analyzer wherein the lexicon and morphological rules are constructed using finite-state devices.

II. Morphology of Kashmiri Language:

Kashmiri, an Indo-Aryan language with roots dating back centuries. Despite its significance, Kashmiri remains relatively underrepresented in computational linguistics research, particularly in the development of NLP tools and resources. One crucial component of NLP systems is morphological analysis, which involves dissecting words into their constituent morphemes to facilitate various language-processing tasks. This paper delves into the development of a morphological analyzer specifically designed for the Kashmiri language, addressing the unique linguistic features and computational challenges involved.

Morphologically, Kashmiri is a mix of both the agglutinating and inflectional type (Koul and Wali, 2009). An agglutinating language consists of polymorphemic words in which each morpheme corresponds to a single lexical meaning or grammatical function. In contrast, inflectional languages do have polymorphemic words but the



lexical meanings and grammatical functions are often fused together. Due to its rich morphological structure, morphological analysis of Kashmiri poses several challenges. Kashmiri shows both agglutinating and inflectional types of morphemic processes in lexical categories such as nouns, verbs, and adjectives. The presence of agglutinative and inflectional affixes, and vocalic and consonantal changes requires robust morphological parsing algorithms which is capable of handling morpheme boundary ambiguity and inflectional variation. Additionally, the absence of explicit vowel markers in the script necessitates the use of probabilistic models and context-based disambiguation techniques to infer the underlying morphological structure of words.

III. Design and Implementation of the Morphological Analyzer:

The development of a morphological analyzer for Kashmiri entails several key steps, including corpus collection and preprocessing, lexicon development, morphological rule formulation, and algorithm implementation. We describe each of these steps in detail below:

Corpus Collection and Preprocessing: The first step in developing the morphological analyzer involves the collection and preprocessing of linguistic data in the form of annotated corpora or text corpora. The corpus serves as the foundation for lexicon construction, rule induction, and algorithm training, and should ideally encompass a diverse range of linguistic phenomena and genres representative of the target language.

Lexicon and Morphological Rules: The lexicon forms the backbone of the morphological analyzer and comprises a comprehensive inventory of lexical entries, morphemes, and their corresponding grammatical attributes. The lexicon is typically derived from the annotated corpus through manual annotation or automatic extraction techniques and serves as the primary source of linguistic knowledge for morphological analysis.

Morphological Parsing Algorithms: Morphological parsing algorithms play a crucial role in segmenting words into morphemes, identifying their grammatical properties, and resolving morph-syntactic ambiguities. In the case of Kashmiri, morphological parsing algorithms must be tailored to handle the language's unique

morphological patterns, including infixation, agglutination, and vowel harmony.

Software Architecture and Implementation Details: The morphological analyzer is implemented as a software module or application capable of accepting input text and producing morphologically analyzed output. The analyzer may be integrated into existing NLP frameworks or deployed as a standalone tool, depending on the intended use case and deployment environment. The implementation of the analyzer involves the integration of lexicons, morphological rules, and parsing algorithms into a cohesive software architecture designed for efficiency, scalability, and ease of maintenance.

Various approaches and techniques are employed in the development of morphological analyzers, including:

- **Rule-Based Approaches:** These approaches utilize linguistic rules and patterns to segment words into morphemes and perform morphological analysis.
- **Finite-State Transducers (FSTs):** FSTs are formal models used to represent morphological rules and perform efficient morphological analysis and generation.
- **Statistical Approaches:** Statistical models and machine learning algorithms are used to learn morphological patterns from large text corpora and make predictions based on statistical probabilities.
- **Hybrid Approaches:** Hybrid models combine rule-based and statistical methods to leverage the strengths of both approaches and improve overall accuracy and performance.

IV. Finite-state Morphology:

Finite state morphology is a widely accepted approach in the computational treatment of morphology. Two-level morphology represents a word as a correspondence between two levels: **Lexical level:** This level concerns the abstract representation of words, focusing on their basic meaning and structure. At the lexical level, morphemes are considered units of meaning. Morphemes are the smallest units of language that carry meaning, such as roots, prefixes, and suffixes. **Surface Level:** The surface level of morphology refers to words' actual form in speech or writing. It includes all the inflections, affixes, and



modifications added to the lexical level to create the final word produced. Surface-level morphology includes variations due to tense, aspect, number, case, and other grammatical features. Morphological parsing is implemented by building mapping rules that map letter sequences such as ‘cups’ on the surface level into morphemes and

Lexical level

| | | | | |
|---|---|---|----|-----|
| C | U | P | +N | +PL |
|---|---|---|----|-----|

Surface level

| | | | | |
|---|---|---|---|--|
| C | U | P | S | |
|---|---|---|---|--|

Fig. 2: Two-Level Morphology

feature sequences like ‘cup + N + PL’ on the lexical level Fig.1 given below shows these two levels for the word cats. The lexical level has the stem for a word followed by the morphological information - N, -PL which tells us that cats are a plural noun.

A. Rule component:

This consists of rules which map the two representations to each other. Each rule is described through a finite-state transducer. A finite state transducer maps between one set of symbols and another, a finite state transducer does this via a finite automaton. Thus, we usually visualize a Finite State Transducer as a two-auxiliary tape automaton that recognizes or generates pairs of strings.

B. Finite-state Transducer:

Of all the finite-state devices, such as finite-state automata and push-down automata finite-state automata, and push-down automata, finite-state morphology mostly uses finite-state transducers (FST). An FST is simply a classical finite-state automaton whose transitions are labeled with pairs, rather than with single symbols e.g., $\Sigma = \{a:a, b:b, a:c, a:\epsilon, e:\epsilon, \dots\}$. It maps one set of symbols to another, via a finite automaton. The Fig. below shows an FST built over the pairs a:A, b:B and c:C.

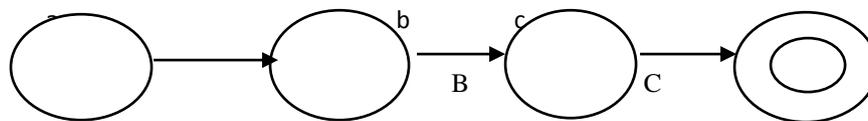


Fig. 3: An Arbitrary Finite-state Transducer

Finite-state devices are theoretically appealing because they are best understood from a mathematical point of view. They are computationally appealing because they are best made for simple, elegant, and highly efficient implementations. Beesley and Karttunen (2003) assert that computing with finite-state devices is attractive because of the following three reasons.

a. First, the mathematical properties of finite-state machines are well understood. This allows one to modify and combine finite-state devices in ways that would be impossible using other traditional algorithmic programs. In other words, this ‘mathematical beauty’ of finite-state devices translates into ‘unparalleled flexibility’, especially

due to properties such as inversion, intersection, union, and composition.

b. Finite-state devices are computationally efficient. Resulting in excellent processing speeds.

c. In most cases, finite-state devices can store a lot of information in relatively little memory (Beesley and Karttunen, 2003)

a. Constructing a Finite-State Transducer:

A finite-state transducer accepts a language stated over pairs of symbols. It consists of states and arcs, where the arcs are labeled by symbol pairs. A set of strings, a language, can be represented by symbols on one side of the arcs of an FST. In Fig. 3, for example, the first string ‘abc’ will be known as the upper language; while the string ‘ABC’ will be



known as the lower language. A set of pairs of strings is called a relation. For example, the relation for the FST given in Fig. 3 is {<'abc', 'ABC'>}. In other words, each path of the transducer represents a pair of strings in the relation. If the upper language of the transducer is the string in the lexical form, and the lower language of the

transducer is the string in the surface form, then the transducer so formed can map between the lexical and surface representations. Consider the example in Fig. 4. It shows a finite-state transducer for an English word walks. Whose root is walk.

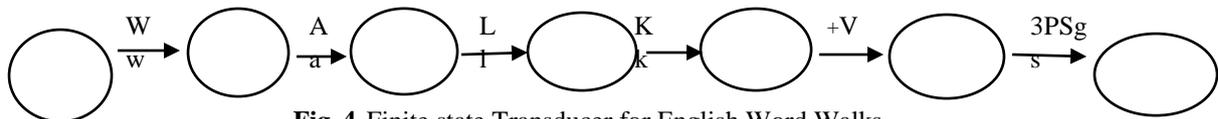


Fig. 4: Finite-state Transducer for English Word Walks

Walks + Verb + 3PSg
w a l k s

using the above transducer, the word walks (surface form). Maps to string “walk+Verb+3PSg” (lexical form), which means:

- The traditional base form is ‘walk’.
- The word ‘walks’ is a verb.
- The word ‘walks’ is in the third person singular form.
- The relation between the surface and lexical forms is {<“walk + Verb + 3PSg”, walks”>}

Notice the new multi-character symbols in Fig.4 +Verb and +3PSg are single symbols, with multi-character Print names. These symbols or tags are chosen and defined by linguists who build the system. It may be noted that FST of Fig. 4 gives the rule to associate lexical and surface representations. When all the lexicons and rules (such as the one described above) are defined and compiled into finite-state transducers, they can be combined together using set operations like union, intersection, and composition (Beesley and Karttunen 2003). This forms a network of transducers. The pairs that the network as a whole accepts are those that are accepted by any one of the component transducers.

b. Morphological Analysis and Generation

Two-level morphology is used as a model for morphological analysis and generation (Koshenniemi, 1997). The term morphological analysis is used for transformation from surface representation to lexical representation (eg. Playing play+Verb+Prog). At some places the term ‘recognizing a word’ is also used to indicate morphological analysis. The opposite of analysis is

generation, i.e., to generate surface strings from lexical strings. It is used in exactly the opposite way from analysis (e.g. walk + Verb + Prog walking). Since transducers are inherently bi-directional, due to the inversion property of finite-state devices, rules written for generation can be used for an analyzer and vice versa. Thus building morphological analyzers/generators requires the same kind of rule formation.

c. Building a Finite-State Morphological Analyzer

To build a finite-state morphological analyzer we need the model of rules (or rule components as shown in Fig. 3). Finite-state morphological analyzer is always composed together with a lexicon. Such a lexicon consists of words represented by FSTs. When all the lexicons and rules are defined by the linguist and compiled into finite-state transducers, they can be combined using any operations like union, intersection, and composition (Beesley and Karttunen, 2003). For some natural languages, it is possible and convenient to divide up the work, doing nouns, verbs, and adjectives separately, the resulting sub-language transducers can then simply be unioned together when they are finished (Beesley and Karttunen, 2003). Most morphological analyzers are equipped with lexicons.

d. → Problems in Rule Formation:

According to Beesley and Karttunen (2003), there are two central problems in morphology as given below. The third problem discussed below has been taken from Sarkar (1993).



a. Word Formation:

Word formation is also called morphotactics or, in other traditions, morph-syntax. The morphemes of a word are constrained to appear in certain combinations and orders. These constraints need to be considered while forming rules.

b. Phonological and orthographical Alternation:

The spelling of a morpheme often depends on the environment. Thus those morphemes that can change shape need to be taken into account. For example; in English note the following alterations (among many others)

- Pity is realized as piti in the context of a following less (E)
- dOd (K) ‘milk’ (E) dOdIvo:l (K) ‘milkman’ (E)
- ga:tal (K) ‘wisdom’ (E) ga:taul (K) ‘wise man’ (E)

As per Beesley and Karttunen (2003), a similar phenomenon appears in almost all languages.

V. Nominal morphology:

Nouns in Kashmiri can be classified according to the traditional classification scheme.

- Proper noun {Human: Animate, non-human: animate and inanimate nouns}
- Common noun {count, mass}

Nouns are not formally distinguished for being definite or indefinite. An optional indefinite marker /-ah/ is used with the nouns. For example;
/asəl kami:zah ha:v/ (show me a good shirt)
good shirt show

/plastik kursijah akh dim/ (give me a plastic chair)
plastic chair one give

/wazul kalmah akh a:n/ (give me a red pen)
red pen one get

Similarly, generic qualities are expressed by the plural form. For example;
/juri tʃʰi suku:l gatsʰa:n/ (children go to school.)
Children aux school going

/kori tʃʰi imoʃnal a:sa:n/. (girls are emotional.)
girls aux emotional are

a. Noun Inflection

Nouns in Kashmiri are inflected for gender, number, and case.

i. Gender:

Nouns, both animate and inanimate, are divided into masculine and feminine. Whereas, animates follow the natural gender system, the gender of a large number of inanimate nouns is often unpredictable. However, gender of a large number of inanimate nouns can be predicted by their endings. The gender formation processes involve, suffixation, changes in vowels and consonants and suppletion. Consider the following examples in the table;

| Masculine | Feminine | Process |
|--------------------------|----------------------------|--------------------------|
| /lədʒki/ boy | /ku:r/ girl | Suppletion |
| /votsʰ/ Calf (male) | /vatsʰir/ Calf (female) | Suffixation |
| /b'o:r/ Cat (male) | /b'ə:r/ Cat (female) | Vowel change |
| /mo:l/ father | /mo:dʒ/ Mother | Final consonant change |
| /kotur/ Pigeon (male) | /kotir/ Pigeon (female) | Penultimate vowel change |

Table 1: Gender Formation Processes



ii. Number:

Kashmiri exhibits a two number system viz., singular and plural. The plural form of nouns are obtained from singular noun by suffixation, palatalization, and change in vowel. Plural nouns may vary with gender, however, no distinction is made between animate and inanimate.

Masculine Plurals

| Singular | Plural | Process |
|--------------------|----------|--------------------------|
| /bə:təl/ Bottle | /bə:tli/ | Suffixation |
| /gagur/ Rat | /gagir/ | Penultimate vowel change |
| /kot/ Boy | /kət/ | Vowel change |
| /dʊn/ walnut | /dun/ | palatalization |

Table 2: Masculine Plurals

Feminine Plural

| Singular | Plural | Process |
|-------------------------|----------|----------------------------------|
| /nər/ Arm | /nari/ | Suffixation |
| /də:r/ Window | /da:ri/ | Suffixation |
| /ku:r/ Girl | /kori/ | Suffixation |
| /zət/ Piece of cloth | /zatʰi/ | Consonant change and suffixation |
| /kutʰ/ Room | /kutsʰi/ | Consonant change and suffixation |
| /vazidʒ/ Red | /vazdʒi/ | Suffixation |

Table 3: Feminine Plurals

iii. Case:

Case suffixes added to nouns and noun phrases occur as bound morphemes. The following table gives the case suffixes added to the nouns agreeing in number and gender

| Case | Masculine | | Feminine | |
|--------------|-----------|----------|----------|----------|
| | Sg | Pl | Sg | Pl |
| Nominative | unmarked | unmarked | unmarked | unmarked |
| Absolutive | unmarked | unmarked | unmarked | unmarked |
| Dative | -as/is | -an | -i | -an |
| Ergative | -an | -av | -i | -av |
| Locative | -as/-is | -an | -i | -av |
| Ablative | -i | -av | -i | -an |
| Instrumental | -i | -av | -i | -av |
| Genitive I | -as | -an | -i | -an |



| | | | | |
|----------|----|-----|----|-----|
| 2 | -i | -av | -i | -av |
| vocative | -a | -av | -i | -av |

Table 4: Case Suffixes

VI. Verb

The verb stems in Kashmiri end in consonants. Only seven verb stems end in vowels. These are /khe-, tʃʰe-, pe-, he-, ni-, di-, and yi-/ (eat, drink, fall, get, take, give and come). Verbs are inflected for person, gender, number, and tense in Kashmiri. Koul and Wali (2006) has classified Kashmiri verbs into the following types:

1. **Intransitive verbs:** Most of these verbs mark their subjects in the nominative case in all the tenses except a few that mark their subjects in ergative case in the past tense. Consider the following examples:

/bi votus/
I reached

/su tʃʰi paka:n/
He is coming

/me nots/
I danced

/tim nots/
He danced

2. **Transitive verbs:** Transitive verbs mark their subject nominative in the non-perfective sentences and ergative in perfective sentences. Direct objects are marked nominative or dative as dictated by the person hierarchy. Consider the following examples:

/beniʃan kʰov ke:k/
Benish(erg) ate a cake

/me pər kita:b/
I read a book

/maʃʃaran lo:j ʃuris/
Teacher beat the student (dat)

/bi dim beniʃas kita:b/
I will give Benish(dat) a book

/beniʃan likʰ tʰesis tre tʃʰapʃar/
Benish (erg) wrote three chapters of her thesis

3. **Causative verbs:** These verbs are derived from the intransitive or transitive verbs. These verbs are transitive in nature. Consider the following examples:

| <i>From Intransitives</i> | <i>From Transitives</i> |
|---|--------------------------------------|
| /ruk-ruka:v-ruka:vna:v/ (stop) | /kʰe-kʰja:v-kʰjavna:v/ (eat) |
| /te:z-te:zra:v-te:zravnav/ (fast) | /tʃe-tʃa:v-tʃavna:v/ (drink) |
| /as-asna:v-asnavna:v/ (laugh) | /par-parina:v-parna:vna:v/ (read) |
| /daz-za:l-za:lina:v-zalnavna:v/ (burn) | /bo:z-bo:zna:v-boznavna:v/ (hear) |

Table 5: Causative Verbs

4. **Ingestive verbs:** These verbs include khe-, tʃe-, vuch-, par- (eat, drink, see, read). These verbs are different from other causatives because they always mark their object as nominative or dative.

5. **Dative verbs:** This group comprises mostly verbs of knowledge, belief, perception and

mental or physical state. These mark their subject dative in all the tenses and aspects. The thematic object is marked nominative and shows agreement with the verb. Consider the following examples:

/banun/ to become
/ba:sun/ to live



/tagun/ to have the ability to do
/jats^hun/ to wish
/jun/ to come
/za:gun/ to keep an eye on

6. **Conjunct verbs:** these verbs are formed by adding a group of verbs to nouns, adjectives and adverbs. This group of verbs forming conjunct verbs with nouns, adjectives and adverbs are called vectors. These include /a:sun/ to be, /karun/ to do, /rozun/ to live, /khasun/ to go up, /dyun/ to give, /jun/ to come, and /pjun/ to fall. The vector verb is placed in the second position and carries the verbal inflection. Consider the following examples:

/sra:n karun/ to bathe
/ja:d karun/ to cram
/ja:d pⁱun/ to remember
/hⁱur k^hasun/ to go upstairs
/dam dⁱun/ to rest
/behit rozun/ to stay idle

7. **Compound verbs:** compound verbs are formed by adding certain auxiliary verbs to the conjunctive participle of the main verb. These auxiliaries are

called explicators or operators. In these verbs the original meaning of the explicator is lost. Koul and Wali (2006) state that these and some aspectual values such as benefaction, intensification or completion of an action. The explicators include the verbs such as /jun, nⁱun, pⁱon, dⁱon, gats^hun, tsalun, t^ha:vun/ (to come, to take, to fall, to give, to go, to flee, to keep).

8. **Modal verbs:** Modality is expressed by the explicators such as /a:sun, pⁱon, pazun, lagun/ (to be, to fall, to be proper, to stick). These explicators take the inflectional markers in a sentence.

6.1. Tense

Kashmiri language makes a distinction between past, present and future tenses. In all these tenses complex case and agreement pattern is visible in the Kashmiri language. The verbs show agreement on the basis of number and gender. The verbs also show a pronominal agreement linked with the nominative, ergative and dative cases.

Inflectional Forms of auxiliary *chu* Agreeing with Nominative Subjects

| Person | Maculine | | Feminine | |
|-----------------|---------------------|----------------------|---------------------|---------------------|
| | Sg | Pl | Sg | Pl |
| 1 st | -t ^h us | -t ^h i | -t ^h as | -t ^h a |
| 2 nd | -t ^h ukh | -t ^h v(i) | -t ^h akh | -t ^h avi |
| 3 rd | -t ^h u | -t ^h i | -t ^h i | -t ^h a |

Table 6: Inflectional Forms of Auxiliary 'chu'

Note that the vocalic part of the inflection refers to the gender number features. The consonantal part represents the person features.

The person features are absent in 1st plural and 3rd person.

| | | | |
|----|----------------|--------------------|--------------|
| 1. | b ⁱ | ts ^h us | para:n |
| | I (NOM) | Aux (am) | Reading (pr) |
| | I am reading. | | |

| | | | |
|----|------------------|--------------------|--------------|
| 2. | ts ⁱ | ts ^h uk | para:n |
| | you (NOM) | aux (are) | Reading (pr) |
| | You are reading. | | |

| | | | |
|----|--------------------|-------------------------------------|--------------|
| 3. | Su/si | Ts ^h u/ts ^h i | para:n |
| | He/she (NOM) | Aux (is) | Reading (pr) |
| | He/she is reading. | | |



6.1.1. Past Tense

Kashmiri language has a morphological mechanism for dividing past into three types which are referred as proximate past (pp), indefinite past (ip) and remote past (rp). Transitive verbs and

certain special intransitives bear suffixes for pp, ip, and rp. Most intransitives do not morphologically distinguish between pp and ip although these are marked for rp.

| Tense | Msg | Mpl | Fsg | Fpl |
|-------|--------|----------|----------|----------|
| Pp | -v | -ji | -ji | -ji |
| Ip | -jo:v | -e:ji | -e:ji | -e:ji |
| Rp | -e:jov | -e:je:ji | -e:je:ji | -e:je:ji |

Table 7: Distinction between Pp, Ip, and Rp in verbs of Kashmiri

In case of the verb stems of transitive, ergative intransitive and special intransitive uniform vocalic and consonantal changes take place. Verbs have been classified into three categories in Kashmiri depending on the conjugation patterns in the past tense (Koul and Wali, 2006).

First Conjugation: Transitive and Ergative Intransitive

Second Conjugation: Special Intransitive

Third Conjugation: Regular Intransitive

First Conjugation

| verb | person | masculine | feminine |
|---------------|--------|---|--|
| /par/ read | first | /par'io:v/ (IP, Sg), /pare:ji/ (IP, Pl) /parejo:v/ (RP, Sg), /pare:je:ji/ (RP, Pl) | /pare:ji/ (IP, Sg), /pare:ji/ (IP, Pl) /pare:je:ji/ (RP, Pl) /pare:je:ji/ (RP, Pl) |
| | Second | /par'io:th/ (IP, Sg) /pareja:th/ (IP, Pl) /pareja:th/ (RP, Sg) /pareje:ja:th/ (RP, Pl) | /pareja:th/ (IP, Sg) /pareja:th/ (IP, Pl) /pareje:ja:th/ (RP, Sg) /pareje:ja:th/ (RP, Pl) |
| | third | /par'io:v/ (IP, Sg), /pare:ji/ (IP, Pl) /parejo:v/ (RP, Sg), /pare:je:ji/ (RP, Pl) | /pare:ji/ (IP, Sg), /pare:ji/ (IP, Pl) /pare:je:ji/ (RP, Pl) /pare:je:ji/ (RP, Pl) |

Table 8: First Conjugation in Kashmiri

Second Conjugation (Special Intransitives)

| verb | person | masculine | feminine |
|--------------------|--------|--|--|
| /rukun/ To stop | first | /ruk'io:s/ (PP, Sg), /ruke:j/ (PP, Pl) /ruke:jo:s/ (IP, Sg), /ruke:je:ji/ (IP, Pl) /ruke:je:jo:s/ (RP, Sg) /ruke:je:je:ji/ (RP, Pl) | /ruke:jas/ (PP, Sg), /ruke:ji/ (PP, Pl) /ruke:je:jas/ (IP, Sg) /ruke:je:ji/ (IP, Pl) /ruke:je:jas/ (RP, Sg) /ruke:je:ji/ (RP, Pl) |
| | Second | /ruk'io:k ^h / (PP, Sg) /ruke:ji/ (PP, Pl) /ruk'io:k ^h / (IP, Sg) | /ruke:jak ^h / (PP, Sg) /ruke:ji/ (PP, Pl) /ruke:jak ^h / (IP, Sg) |



| | | | |
|--|-------|--|---|
| | | /ruke:je:ji/ (IP, Pl) /ruke:je:ja:kʰ/ (RP, Sg) /ruke:je:jivi/ (RP, Pl) | /ruke:javi/ (IP, Pl) /ruke:je:jakh/(RP, Sg) /ruke:je:javi/(RP, Pl) |
| | third | /ruk'o:v/ (PP, Sg) /ruke:ji/ (PP, Pl) /ruke:jo:v/ (IP, Sg) /ruke:je:ji/ (IP, Pl) /ruke:jejo:v/ (RP, Sg) /ruke:je:ji/ (RP, Pl) | /ruke:j/ (PP, Sg) /ruke:ji/ (PP, Pl) /ruke:je:ji/ (IP, Sg) /ruke:je:ji/ (IP, Pl) /ruke:je:ji/ (RP, Sg) /ruke:je:ji/ (RP, Sg) |

Table 9: Second Conjugation in Kashmiri

Third Conjugation

| verb | person | masculine | feminine |
|--------------|--------|--|--|
| /ji/ come | first | /a:ja:s/ (IP, Sg), /a:je:ji/ (IP, Pl) /a:je:ja:s/ (RP, Sg) /a:je:ji/ (RP, Pl) | /a:je:jas/ (IP, Sg), /a:je:ji/ (IP, Pl) /a:je:ja:s/ (RP, Sg) /a:je:ji/ (RP, Pl) |
| | Second | /a:ja:kʰ / (IP, Sg) /a:je:jvi/ (IP, Pl) /a:je:jakʰ / (RP, Sg) /a:je:jvi/ (RP, Pl) | /a:je:jakʰ / (IP, Sg) /a:je:jvi/ (IP, Pl) /a:je:jakʰ RP, Sg) /a:je:jvi/(RP, Pl) |
| | third | /a:jo:v/ (IP, Sg) /a:je:ji/ (IP, Pl) /a:je:jo:v/ (RP, Sg) /a:je:ji/ (RP, Pl) | /a:je:ji/ (IP, Pl) /a:je:ji/ (IP, Pl) /a:je:ja:ji/ (RP, Sg) /a:je:ji/ (RP, Sg) |

Table 10: Third conjugation in Kashmiri

‘Transitive ergative subjects in the first and third person do not affect the agreement. Second person ergative subjects mark the verb with suffixes (-th/-ov (sg/pl)). Honorific forms employ -vi suffix which is added to first and third person forms’ (Wani, 2013).

Ergative intransitives are neutral with respect to first and third persons but inflect for second person. Verbs of the second conjugation consist of nearly sixty-seven intransitive verbs (Koul and Wali, 2009). These verb shows agreement with nominative subject and undergo various phonemic changes like verbs of the first conjugation. Verbs of the third conjugation show agreement with the nominative subject and include most intransitive

verbs. These verbs do not distinguish between indefinite past and proximate past morphologically.

6.1.2. Future Tense

Future tense is marked by suffixation in Kashmiri. But, the suffixes do not show any gender distinction. These suffixes have been categorized into two types depending upon their agreement patterns.

Type I suffixes: These suffixes agree with the nominative subjects and include verbs such as *kʰe* ‘eat’, *ji* ‘come’, *di* ‘give’, *par* ‘read’, *tʰal* ‘wash’. These suffixes differ on the basis of the sound i.e., vowel of consonant at the end of the verb stems. These suffixes also agree with the number and person of the subject. Consider the following tables:



| Person | Singular | Plural |
|--------|-----------------|--------|
| First | -mi | -mav |
| Second | -k ^h | -jiv |
| Third | -ji | -jan |

Table 11: Suffixes for verb stems ending in a vowel sound

| Person | Singular | Plural |
|--------|------------------|--------|
| First | -i | -av |
| Second | -ak ^h | -iv |
| Third | -i | -an |

Table 12: Suffixes for the verb stems ending in a consonant sound

Type II suffixes are those suffixes which are added to transitive verbs only, and show a complex agreement with both the subjects and objects in a complex manner. These suffixes are determined by inanimate direct objects and pronominal objects.

| Person | Singular | Plural |
|--------|----------|--------|
| First | -ah | -iho |
| second | -ih | -iha |
| third | -j | -in |

Table 13: Verb suffixes

Subject+ inanimate direct objects

| Person | Object | |
|-----------------|----------|---------------------|
| | Singular | Plural |
| First singular | -an | -ak ^h |
| First plural | -iho:n | -iho:k ^h |
| Second singular | -iha:n | -iha:k ^h |
| Second plural | -ihu:n | -ihu:k ^h |
| Third singular | -jas | -jak ^h |
| Third plural | -inas | -inak ^h |

Table 14: Verb suffixes with respect to inanimate direct objects

Examples

1. /bi k^hema:n ji tsu:n^h/ I will eat this apple.
I eat-FUT this apple
2. /bi k^hemak^h jim tsu:n^h/ I will eat these apples
I eat-FUT-pl these fruits

VII. Applications and Future Directions

The morphological analyzer for Kashmiri holds significant potential for various applications in natural language processing, including part-of-speech tagging, named entity recognition, sentiment analysis, and machine translation. By providing insights into the grammatical structure and semantic composition of words, the analyzer enables more accurate and contextually aware

language processing across a wide range of domains and applications.

Furthermore, the morphological analyzer serves as a foundational component for the development of higher-level language processing tools and resources, such as syntactic parsers, semantic analyzers, and language generation models. Integration with existing NLP frameworks and platforms facilitates the adoption and deployment of Kashmiri language processing tools in real-world applications, thereby empowering users to interact with digital content and services in their native language.

Future directions for improving Kashmiri language processing tools include the expansion of the lexicon to encompass additional dialectal



variants and lexical forms, the refinement of morphological rules to handle rare and irregular morphological patterns, and the integration of machine learning techniques for automatic rule induction and morphological disambiguation. Collaboration with linguistic experts and native speakers can further enhance the quality and coverage of linguistic resources and contribute to the development of more robust and linguistically informed language processing tools for Kashmiri.

VIII. Conclusion:

In conclusion, this paper has presented the design, implementation, and evaluation of a morphological analyzer tailored for the Kashmiri language. We have discussed the linguistic characteristics of Kashmiri, the methodologies employed in morphological analysis, and the technical intricacies involved in developing the analyzer. Through rigorous evaluation and analysis, we have demonstrated the efficacy and potential applications of the morphological analyzer in various natural language processing tasks.

Moving forward, efforts to improve Kashmiri language processing tools should focus on expanding linguistic resources, refining parsing algorithms, and fostering collaboration between researchers, linguists, and native speakers. By leveraging advances in computational linguistics and machine learning, we can pave the way for the development of more sophisticated and contextually aware language processing tools that cater to the diverse linguistic needs and preferences of Kashmiri speakers.

References:

- [1]. Beesley, Kenneth R., and Lauri Karttunen. 2003. 'Finite-state morphology: Xerox tools and techniques.' CSLI, Stanford, 359-375.
- [2]. Koul, O.N. and Wali, K. 2006. Modern Kashmiri grammar. Dunwoody Press Springfield.
- [3]. Wani, Sajad H. 2013. Kashmiri to English Machine Translation: A Study of Morpho-Syntactic Divergence Patterns, (PhD Thesis), University of Kashmir.