Interdisciplinary Journal of Linguistics
Volume [11] 2018, Pp.54-66

Developing a Morphological Analyser/Generator for Kashmiri

Nazima Mehdi[*]
Aadil Amin Kak[*]

Abstract

Morphological Analyser/Generator is an essential component of any NLP application. A morphological analyzer or generator supplies information concerning morphosyntactic properties of the words it analyses or constructs. Morphological Analysis and Generation are important components for building computational grammars as well as Machine Translation (MT). Morphological Analyzer is a program for analyzing the morphology of an input word; the analyzer reads the inflected surface form of each word in a text and provides its lexical form while Generation is the inverse process. Both Analysis and Generation make use of lexicon. A Morphological analyser is developed using either a paradigm based approach or a rule based approach. Words can be categorized into paradigms based on the similarity in grammatical features and word formation process. It is observed that for a particular grammatical feature there are some similarities in word formation process. A Paradigm defines all word forms that can be generated from given stem along with grammatical feature set associated with each word form. For paradigm construction, sample corpus of inflected words is taken.

This paper deals with the development of a Kashmiri morphological analyser/generator based on a paradigm based approach. Different paradigms are built for nouns, verbs, adjectives, and adverbs taking all the possible inflections into account. A list is prepared for all possible word forms from the lexicon. The words which take similar set of suffixes are grouped into single paradigm. For this analysis the linguistic resources required by the morphological analyzer include a lexicon and inflection rules for all paradigms. Inflection rules include the inflectional suffixes, and internal vowel change to the root to get its inflected forms.

Keywords: Morphological Analyser, NLP, MT, Paradigm, Lexicon

1. Introduction

Morphological Analysis

Morphological analysis is the process of providing grammatical information about the word on the basis of properties of the morpheme it contains. It is an integral part of the larger natural language processing projects such as text to speech synthesis, information extraction and machine translation.

A morphological analyzer is the program for analyzing the morphology of an input word. The analyzer includes the recognition engine, identifying suffixes, and finding a stem within the input word algorithms (Shrivastava et al., 2005). A morphological analyzer takes a complete word form and the syntactic and morphological properties of the word as its input. Morphological analyzers are composed of three parts (Kumar 2013):

- Morpheme lexeme
- Set of rules governing the spelling and composition of morphologically complex words.
- Decision algorithm

[*] Department of Linguistics, University of Kashmir

Morphological analyzer and generator are the two essential and basic tools for building any natural language processing application. It supplies information concerning morphosyntactic properties of the words it analyses or constructs. Morphological analyzer is a program for analyzing the morphology of an input word, the analyzer reads the inflected surface form of each word in a text and provides its lexical form, like for nouns it will provide gender, number, and case information, likewise for verbs it will provide tense, aspect and modularity. Whereas generation is the inverse process i.e., given a root and its grammatical features it will generate the word forms of the root word.

1.1. Various Methods of Morphological Analysis

Various NLP (Natural Language Processing) research groups have developed different methods and algorithms for morphological analysis. Some of the algorithms are language dependent and some of them are language independent. A brief description of various methods involved in morphological analysis include the following.

- Finite State Automata (FSA).
- Two Level Morphology
- Finite State Transducer (FST).
- Stemmer Algorithm.
- Corpus Based Approach
- DAWG (Directed Acrylic Word Graph)
- Paradigm Based Approach.

The present work is aimed at devising a rule based frame work for morphological analyzer/generator for Kashmiri. The methodology used for the present research is purely qualitative involving paradigm construction. Based on these paradigms, rules are formulated which in-turn will form the base for building a rule based morphological analyzer.

For the present study written corpus was taken as a primary source. The corpus includes literary texts (including all domains) like newspapers, text books, magazines, etc. For this work various Dictionaries were also consulted. Apart from written corpus, e-corpus of 1,000,000 words was also used as an input for the POS tagger for lexicon building.

The design of the morphological analyzer is divided into two main components: Lexicon and Paradigms. Lexicon in its most general sense is synonymous with vocabulary. Lexicon has a special status in generative grammar, where it refers to the component containing all the information about the structural properties of the lexical items in a language (Crystal, 2008: 278). The lexicon is not regarded merely as a long list of words. Rather, it is conceived as a set of lexical resources, including the morphemes of the languages, plus the processes available in the language for constructing words from those resources.

Words can be categorized into paradigms based on the similarity in grammatical features and word formation process. It is observed that for a particular grammatical feature there are some similarities in word formation process. A Paradigm defines all word forms that can be generated from a given stem along with the grammatical feature set associated with each word form. The methodology for developing a morphological analyzer and generator begins by identifying and defining parameters for Kashmiri, which was done in two phases viz,

I. Identification of Different Parts of Speech in the Lexicon

The main parameter for developing a morphological analyzer/ generator for Kashmiri required a Lexicon. The lexicon of a language is its vocabulary or in other words it is an explicit list of every word of the language. It is a cumbersome task to tag every word in a language so POS tagger (Trigrams'n' Tags) was used to tag parts of speech like Noun, Verbs, Adjectives, Pronouns, Postpositions, and Adverbs. Trigrams'n' Tags (TnT) which uses a statistical

approach, is a very efficient statistical part-of-speech tagger and is trainable on different languages. The component for parameter generation trains on tagged corpora. The system incorporates several methods of smoothing and handling unknown words. TnT is not optimized for any particular language. Instead, it is optimized for training on a large variety of corpora and it is very easy to adapt the tagger to a new language, new domain or new tagset and these positive features strongly favored the use of TnT for the present work. Moreover TnT is optimized for speed. For part-of-speech tagging TnT uses second order Markov Model. An important characteristic of TnT tagger is that it not only assign tags to words but also to the probabilities. For the present work TnT is trained using a hierarchical tagset. Tagger was used for assigning a tag to each word in the corpus, implementing the tag-set and tagging-scheme in a tag-assignment algorithm. In other words we can say the output of the tagger consists of POS-tagged files, containing all possible tags for each token. Tagging process requires two files, one is the model parameter for lexical and contextual frequencies and the second one is the input file in the desired format which can be tagged or untagged. If it is a tagged file then the tagger will use only the first column in each line and the rest is stripped. Rest of the line can contain comments, tags or any other material without having any effect on the tagging process. And, if the file is untagged then there will be only one column, and the tagger will use that only for tagging. The diagram given below shows the tagging process.

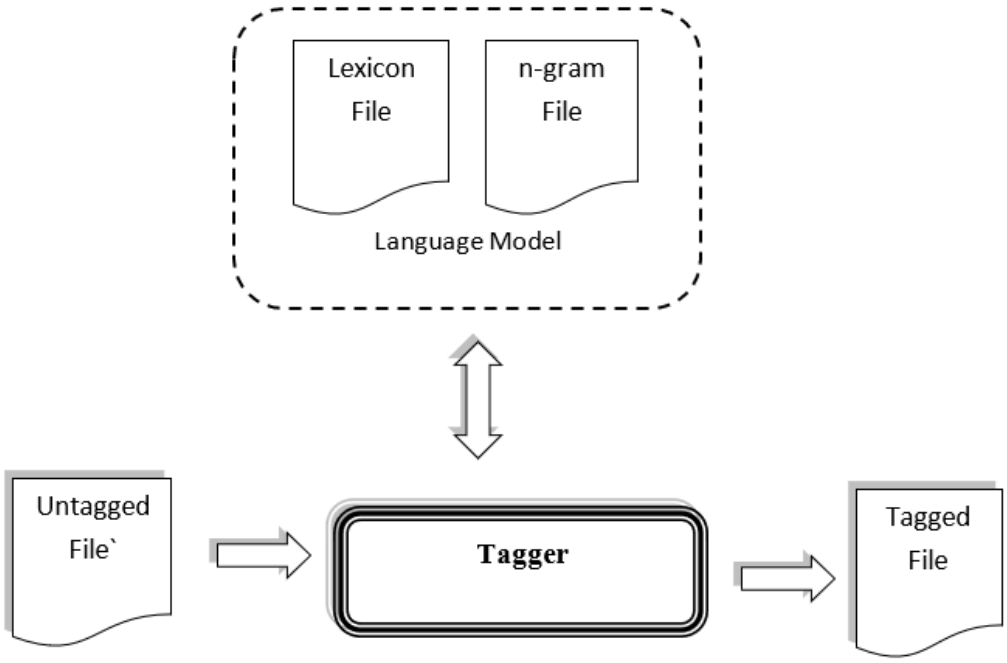

Fig.1.1: Tagging of untagged files.

II. Building Paradigms for POS

For the present research POSs were categorized as declinables and indeclinables on the basis of inflections they take:

- Declinables include all those POS which inflect for Tense, Number, Gender, and Case and include Nouns, Pronouns, Verbs, Adjectives.
- Indeclinables do not show any inflection for Tense, Number, Gender, and Case.

On the basis of the above classification, paradigms were built for Nouns, Verbs, Adjectives, Pronouns, Postpositions and Adverbs on the basis of morphophonemic changes. A paradigm was built for a particular word class taking all the possible features into consideration which gave rise to different word forms of that particular word class, e.g a paradigm for Noun */kul/* 'tree' was built considering all its the grammatical features into consideration and the words which share the same features as that of */kul/* 'tree' paradigm will fall in this paradigm. Likewise, other POSs were handled considering all the grammatical features or morphophonemic changes. Given below are some examples of construction and analysis of such paradigms.

2. Analysis of Nouns

In Kashmiri Nouns are inflected for Gender, Number and Case. An analysis of nouns is given under in the form of paradigms.

2.1 Paradigm for Noun کُل/kul/ 'tree'

Feature description	Possession	کُل /kul/	Analysis
Nom. Sg.	All	کُل /kul/	No change is seen.
Nom. Pl.	All	کُلۍ / kulʲ[1]/	Ultimate consonant gets palatalized (pluralized).
Dat. Sg.	All	کُلِس /kul-is/	/-is/ is added at the word final position.
Dat. Pl.	All	کُلِین /kulʲ-en/	Ultimate consonant is palatalized and/-en/ is added at the word final position.
Erg. Sg.	All	کُلۍ / kulʲ/	Ultimate consonant gets palatalized.
Erg. Pl.	All	کُلیو /kulʲ-ev/	Ultimate consonant gets palatalized /-ev/ is added at the word end.
Gen. Sg	Mas Sg	کُلُیک /kulʲ-uk/	Ultimate consonant gets palatalized /-uk/ is added at the word final position.
	Fem Sg	کُلِچ /kul-iʧ/	/-iʧ/ is added at the word final position.
	Mas Pl	کُلِکۍ / kul-ikʲ/	/-ikʲ/ is added at the word final position.
	Fem Pl	کُلۍ چہِ /kulʲ-ʧi/	Ultimate consonant gets palatalized /-ʧi/ is added at the word final position.
Gen. Pl	Mas Sg	کُلِین ہُنْد /kulʲ-en-hund/	Ultimate consonant gets palatalized. /-en/ is added at the word final position. Suffix /-hund/ is added at the word final position.
	Fem Sg	کُلِین ہِنْز /kulʲ-en-hund/	Ultimate consonant gets palatalized. /-en/ is added at the word final position. Suffix /-hund/ is added at the word final position.

[1] /ʲ/ palatalization at the end is used to denote pluralization as well as ergativity.

	Mas Pl	کُلِین ہِنْدِ /kulʲen-hɨndʲ/	Ultimate consonant gets palatalized. /-en/ is added at the word final position. Suffix /-hɨndʲ/ is added at the word final position.
	Fem Pl	کُلِین ہِنْز /kulʲ-en-hɨnz/	Ultimate consonant gets palatalized. /-en/ is added at the word final position. Suffix /-hɨnzɨ/ is added at the word final position.

3. Analysis of Verbs

Traditionally verbs are classified as intransitive, transitive, causative, dative, conjunct or compound. Finite verbs inflect for tense, aspect, mood and voice, as well as for person, number and gender features of their arguments.

3.1 Paradigm Construction

For the present analysis imperative verbs have been taken into consideration to built paradigms. Paradigm construction for such verbs is given under.

3.1.Paradigm for Verb اَن /an/ 'bring'

<table>
<tr><th>Tense</th><th>Person</th><th colspan="2">Sub</th><th>Obj Mas Sg</th><th>Obj Fem Sg</th><th>Obj Mas Pl</th><th>Obj Fem Pl</th></tr>
<tr><td rowspan="10">Pst</td><td rowspan="10">Ist</td><td>Sg</td><td>Mas</td><td>اوّن
/on/</td><td>اُنہِ
/ənʲ/</td><td>اُنہِ
/ənʲ/</td><td>اَنہِ
/an-i/</td></tr>
<tr><td colspan="2">/me/</td></tr>
<tr><td colspan="6">For obj(mas sg) /a/ → /o/.
For obj(fem sg /mas pl) /a/→/ə/ and final consonant is palatalized.
For obj(fem pl) /-i/ is added at the word final position.</td></tr>
<tr><td>Pl</td><td>Mas</td><td rowspan="2">اوّن
/on/</td><td rowspan="2">اُنہِ
/ənʲ/</td><td rowspan="2">اُنہِ
/ənʲ/</td><td rowspan="2">اَنہِ
/an-i/</td></tr>
<tr><td colspan="2">/asi/</td></tr>
<tr><td colspan="6">For obj(mas sg) /a/ → /o/
For obj(fem sg /mas pl) /a/→/ə/ and final consonant is palatalized.
For obj(fem pl) /-i/ is added at the word final position.</td></tr>
<tr><td>Sg</td><td>Fem</td><td rowspan="2">اوّن
/on/</td><td rowspan="2">اُنہِ
/ənʲ/</td><td rowspan="2">اُنہِ
/ənʲ/</td><td rowspan="2">اَنہِ
/an-i/</td></tr>
<tr><td colspan="2">/me/</td></tr>
<tr><td colspan="6">For obj(mas sg) /a/ → /o/
For obj (fem sg / mas pl) /a/→/ə/ and final consonant is palatalized.
For obj(fem pl) /-i/ is added at the word final position.</td></tr>
<tr><td>Pl</td><td>Fem</td><td rowspan="2">اوّن
/on/</td><td rowspan="2">اُنہِ
/ənʲ/</td><td rowspan="2">اُنہِ
/ənʲ/</td><td rowspan="2">اَنہِ
/an-i/</td></tr>
<tr><td colspan="2">/asi/</td></tr>
</table>

<table>
<tr><td colspan="2"></td><td colspan="6">For obj(mas sg) /a/ → /o/ object.
For obj(fem sg/mas pl) /a/→/ə/ and final consonant is palatalized.
For obj (fem pl) /-i/ is added at the word final position.</td></tr>
<tr><td rowspan="8">Pst</td><td rowspan="8">IInd</td><td>Sg</td><td>Mas</td><td>اوّنُتھ</td><td>اُنِتھ</td><td>اُنِتھ</td><td>اَنیتھ</td></tr>
<tr><td colspan="2">/tse/</td><td>/on-ut^h/</td><td>/ən-it^h/</td><td>/ən-it^h/</td><td>/an^j-et^h/</td></tr>
<tr><td colspan="6">For obj(mas sg) /a /→ /o/ and /-ut^h/ is added at the word final position.
For obj (fem sg/mas pl) /a/→ /ə/ and /-it^h/ is added at the word final position .
For obj(fem pl) ultimate consonant gets palatalized and /-et^h/ is used at the word final position.</td></tr>
<tr><td>Pl</td><td>Mas</td><td>اوّنُو</td><td>اُنِو</td><td>اُنِو</td><td>اَنیو</td></tr>
<tr><td colspan="2">/təhi/</td><td>/on-uv/</td><td>/ən-iv/</td><td>/ən-iv/</td><td>/an^j-ev/</td></tr>
<tr><td colspan="6">For obj(mas sg) /a/→ /o / and /-uv/ is added at the word final position.
For obj(fem sg/ mas pl) /a/→ /ə/ and /-iv/ is added at the word final position.
For obj(fem pl) ultimate consonant gets palatalized and /-ev/ is used at the word final position.</td></tr>
<tr><td>Sg</td><td>Fem</td><td>اوّنُتھ</td><td>اُنِتھ</td><td>اُنِتھ</td><td>اَنیتھ</td></tr>
<tr><td colspan="2">/tse/</td><td>on-ut^h/</td><td>/ən-it^h/</td><td>/ən-it^h/</td><td>/an^j-et^h/</td></tr>
<tr><td colspan="6">For obj(mas sg) /a /→ /o/ and /-ut^h/ is added at the word final position.
For obj (fem sg / mas pl) /a/→/ə/ and /-it^h/ is added at the word final position.
For obj(fem pl) ultimate consonant gets palatalized and /–et^h/ is added at the word final position.</td></tr>
<tr><td>Pl</td><td>Fem</td><td>اوّنُو</td><td>اُنِو</td><td>اُنِو</td><td>اَنیو</td></tr>
<tr><td colspan="2">/təhi/</td><td>/on-uv/</td><td>/ən-iv/</td><td>/ən-iv/</td><td>/an^j-ev/</td></tr>
<tr><td colspan="2"></td><td colspan="6">For obj(mas sg)/a/ →/o / and /-uv/ is added at the word final position.
For obj(fem sg/ mas pl) /a/→/ə/ and /-iv/ is added at the word final position.
For obj(fem pl) ultimate consonant gets palatalized and /-ev/ is added at the word final position .</td></tr>
<tr><td rowspan="4">Pst</td><td rowspan="4">IIIrd</td><td>Sg</td><td>Mas</td><td>اوّن</td><td>اُنیِ</td><td>اُنیِ</td><td>اَنہِ</td></tr>
<tr><td colspan="2">/təm^j/</td><td>/on/</td><td>/ən^j/</td><td>/ən^j/</td><td>/an-i/</td></tr>
<tr><td colspan="6">For obj(mas sg) /a/ → /o/.
For obj(fem sg /mas pl) /a/→/ə/ and final consonant is palatalized.
For obj(fem pl) /-i/ is added at the word final position.</td></tr>
<tr><td>Pl</td><td>Mas</td><td>اوّن</td><td>اُنیِ</td><td>اُنیِ</td><td>اَنہِ</td></tr>
<tr><td colspan="2">/timav/</td><td>/on/</td><td>/ən^j/</td><td>/ən^j/</td><td>/an-i/</td></tr>
</table>

<table>
<tr><td></td><td></td><td colspan="6">For obj(mas sg) /a/ → /o/.
For obj(fem sg / mas pl) /a/→/ə/ and final consonant is palatalized.
For obj(fem pl) /-i/ is added at the word final position.</td></tr>
<tr><td></td><td></td><td colspan="2">Sg Fem
/təmʲ/</td><td>اوّن
/on/</td><td>اُنہِ
/ənʲ/</td><td>اُنہِ
/ənʲ/</td><td>اَنہِ
/an-i/</td></tr>
<tr><td></td><td></td><td colspan="6">For obj(mas sg) /a/ → /o/.
For obj (fem sg/mas pl) /a/→/ə/ and final consonant is palatalized.
For obj (fem pl) /-i/ is added at the word final position.</td></tr>
<tr><td></td><td></td><td colspan="2">Pl Fem
/timav/</td><td>اوّن
/on/</td><td>اُنہِ
/ənʲ/</td><td>اُنہِ
/ənʲ/</td><td>اَنہِ
/an-i/</td></tr>
<tr><td></td><td></td><td colspan="6">For obj(mas sg) /a/ → /o/.
For obj(fem sg / mas pl) /a/→/ə/ and final consonant is palatalized.
For obj(fem pl) /-i/ is added at the word final position.</td></tr>
<tr><td rowspan="8">Prt</td><td rowspan="8">Ist</td><td>Sg Mas
/bɨ ʧʰus/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td>Pl Mas
/əsʲ ʧʰi/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/anaːn/</td></tr>
<tr><td>Sg Fem
/bɨ ʧʰas/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td>Pl Fem
/əsʲ ʧʰi/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td>Sg Mas
/tsɨ ʧʰukʰ/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td>Pl Mas
/tuhʲ ʧʰiv/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td>Sg Fem
/tsɨ ʧʰakʰ/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td>Pl Fem
/tuhʲ ʧʰavɨ/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td rowspan="8">Prt</td><td rowspan="8">IIIrd</td><td>Sg Mas
/su ʧʰu/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td>Pl Mas
/tim ʧʰi/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td>Sg Fem
/sɔ ʧʰa/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td>Pl Fem
/tim ʧʰa/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td>اَنان
/an-aːn/</td><td colspan="2">اَنان
/an-aːn/</td></tr>
<tr><td></td><td colspan="7">Across all the forms of Present tense verb ‘be’ is followed by the main verb.</td></tr>
</table>

		/-a:n/ is added at the word final position.				
Fut	I[st]	Sg Mas /bɨ/	اَنہٖ /an-ɨ/	اَنہٖ /an-ɨ/	اَنہٖ /an-ɨ/	اَنہٖ /an-ɨ/
		/-ɨ/ is added at the word final position.				
		Pl Mas /əsʲ/	اَنَو /an-av/	اَنَو /an-av/	اَنَو /an-av/	اَنَو /an-av/
		/-av/ is added at the word final position.				
		Sg Fem /bɨ/	اَنہٖ /an-ɨ/	اَنہٖ /an-ɨ/	اَنہٖ /an-ɨ/	اَنہٖ /an-ɨ/
		/-ɨ/ is added at the word final position.				
		Pl Fem /əsʲ/	اَنو /an-av/	اَنو /an-av/	اَنو /an-av/	اَنو /an-av/
		/-av/ is added at the word final position.				
Fut	II[nd]	Sg Mas /tsɨ/	اَنَکھ /an-akʰ/	اَنَکھ /an-akʰ/	اَنَکھ /an-akʰ/	اَنَکھ /an-akʰ/
		/-akʰ/ is added at the word final position.				
		Pl Mas /tuhʲ/	اَنِو /ən-iv/	اَنِو /ən-iv/	اَنِو /ən-iv/	اَنِو /ən-iv/
		/a/ →/ə/ /-iv/ is added at the word final position.				
		Sg Fem /tsɨ/	اَنَکھ /an-akʰ/	اَنَکھ /an-akʰ/	اَنَکھ /an-akʰ/	اَنَکھ /an-akʰ/
		/-akʰ/ is added at the word final position.				
		Pl Fem /tuhʲ/	اَنِو /ən-iv/	اَنِو /ən-iv/	اَنِو /ən-iv/	اَنِو /ən-iv/
		/a/→/ə/ /- iv / is added at the word final position.				
Fut	III[rd]	Sg Mas /su/	اَنہٖ /an-i/	اَنہٖ /an-i/	اَنہٖ /an-i/	اَنہٖ /an-i/
		/-i/ is added at the word final position.				
		Pl Mas /tim/	اَنن /an-an/	اَنن /an-an/	اَنن /an-an/	اَنن /an-an/
		/-an/ is added at the word final position.				
		Sg Fem /sɔ/	اَنہٖ /an-i/	اَنہٖ /an-i/	اَنہٖ /an-i/	اَنہٖ /an-i/
		/-i/ is added at the word final position.				

		Pl	Fem	اَنَن /an-an/	اَنَن /an-an/	اَنَن /an-an/	اَنَن /an-an/
		/tim/					
		/-an/ is added at the word final position.					

4. Analysis of Adjectives

Adjectives are those words that modify Nouns. They act as complements of verbs (Copula) and modify degree adverbs. In Kashmiri, adjectives decline for number, gender, and case of nouns they modify. There are two types of adjectives: (i) Simple, and (ii) Derived. Derived adjectives are formed from nominal, verbal, and other adjectival bases by adding certain suffixes.

4.1 Paradigm construction

A paradigm representation of Adjectives is given below.

4.1 Paradigm for Adjective لوو /lo:v/ 'soft'

Feature Description	Possession	لوو /lo:v/	Analysis
Nom. Mas. Sg.	All	لوو /lo:v/	No change is seen.
Nom. Mas. Pl.	All	لأوٕ /ləːvʲ/	/o:/ → /əː/ Ultimate consonant is palatalized.
Nom. Fem. Sg.	All	لأو /ləːv/	/o:/ →/əː/
Nom. Fem. Pl.	All	لاوِ /laːv-i/	/o:/ → /a:/ /-i/ added at the word final position.
Erg. Mas. Sg.	All	لأوٕ /ləːvʲ/	/o:/→ /əː/ Ultimate consonant is palatalized.
Erg. Mas. Pl.	All	لاوِیو /laːvʲ-ev/	/o:/ → /a:/ Ultimate consonant is palatalized /-ev/ is added at the word final position.
Erg. Fem. Sg.	All	لاوِ /laːv-i/	/o:/ → /a:/ /-i/ is added at the word final position.
Erg. Fem. Pl.	All	لاوِیو /laːvʲ-ev/	/o:/ → /a:/ Ultimate consonant is palatalized. /-ev/ is added at the word final position.
Abl. Mas. Sg.	All	لاوِ /laːv-i/	/o:/ → /a:/ /-i/ is added at the word final position.
Abl. Mas. Pl.	All	لاوِیو /laːvʲ-ev/	/o:/ → /a:/ Ultimate consonant is palatalized. /-ev/ is added at the word final position.
Abl. Fem. Sg.	All	لاوِ /laːv-i/	/o:/ → /a:/ /-i/ is added at the word final position.
Abl. Fem. Pl.	All	لاوِیو /laːvʲ-ev/	/o:/ → /a:/ Ultimate consonant is palatalized. /-ev/ is added at the word final

			position.
Dat. Mas. Sg.	All	لأوس /ləːv-is/	/oː/ →/ əː/ /-is/ is added at the word final position.
Dat. Mas. Pl.	All	لاوِن /laːvʲ-en/	/oː/ → /aː/ Ultimate consonant is palatalized /-en/ is added at the word final position.
Dat. Fem. Sg.	All	لاوِ /laːv-i/	/oː/ → /aː/ /-i/ is added at the word final position.
Dat. Fem. Pl.	All	لاوِن /laːvʲ-en/	/oː/ → /aː/ Ultimate consonant is palatalized /-en/ is added at the word final position.
Gen. Mas. Sg.	All	لوو /loːv/	No change is seen.
Gen. Mas. Pl.	All	لاوِن /laːvʲ-en/	/oː/ → /aː/ Ultimate consonant is palatalized. /-en/ is added at the word final position.
Gen. Fem. Sg.	All	لاوِ /laːv-i/	/oː/ → /aː/ /-i/ is added at the word final position.
Gen. Fem Pl.	All	لاوِن /laːvʲ-en/	/oː/ → /aː/ Ultimate consonant is palatalized. /-en/ is added at the word final position.
Gen.Mas.Sg.	Mas Sg	لأوِ سُنْد /ləːvʲ-sund/	/oː/ → /aː/ Ultimate consonant is palatalized. Suffix /-sund/ is added at the word final position.
	Fem Sg	لأوِ سِنْز /ləːvʲ-sɨnz/	/oː/ → /aː/ Ultimate consonant is palatalized. Suffix /-sɨnz/ is added at the word final position.
	Mas Pl	لأوِ سِنْدِ /ləːvʲ-sɨndʲ/	/oː/ → /aː/ Ultimate consonant is palatalized. Suffix /-sɨndʲ/ is added at the word final position.
	Fem Pl	لأوِ سِنْزِ /ləːvʲ-sɨnzɨ/	/oː/ → /aː/ Ultimate consonant is palatalized. Suffix /-sɨnzɨ/ is added at the word final position.
Gen.Mas.Pl	Mas. Sg	لاوِن هُنْد /laːvʲ-en-hund/	Ultimate consonant is palatalized. /-en/ is added at the word final position. Suffix /-hund/ is added at the word final position.

	Fem Sg	لاوِيِن ہِنْز /la:vʲ-en-hɨnz/	Ultimate consonant is palatalized. /-en/ is added at the word final position. Suffix /-hɨnz/ is added at the word final position.
	Mas Pl	لاوِين ہِنْدِ /la:vʲ-en-hɨndʲ/	Ultimate consonant is palatalized. /-en/ is added at the word final position. Suffix /-hɨndʲ/ is added at the word final position.
	Fem Pl	لاوِين ہِنْزِ /la:vʲ-en-hɨnzɨ/	Ultimate consonant is palatalized. /-en/ is added at the word final position. Suffix /-hɨnzi/ is added at the word final position.
Gen.Fem.Sg	Mas. Sg.	لاوِ ہُنْد /la:v-i-hund/	/o:/→/a:/ /-i/ is added at the word final position. Suffix /-hund/ is added at the word final position.
	Fem. Sg.	لاوِ ہِنْز /la:v-i-hɨnz/	/o:/→/a:/ /-i/ is added at the word final position. Suffix /-hinz/ is added at the word final position.
	Mas. Pl.	لاوِ ہِنْدِ /la:v-i-hɨndʲ/	/o:/→/a:/ /-i/ is added at the word final position. Suffix /-hɨndʲ/ is added at the word final position.
	Fem. Pl.	لاوِ ہِنْزِ /la:v-i-hɨnzɨ/	/o:/→/a:/ /-i/ is added at the word final position. Suffix /-hɨnzɨ/ is added at the word final position.
Gen.Fem.Pl	Mas. Sg	لاوِين ہُنْد /la:vʲ-en-hund/	Ultimate consonant is palatalized. /-en/ is added at the word final position. Suffix /-hund/ is added at the word final position.
	Fem Sg	لاوِين ہِنْز /la:vʲ-en-hɨnz/	Ultimate consonant is palatalized. /-en/ is added at the word final position. Suffix /-hɨnz/ is added at the word final position.

	Mas Pl	لاوِن ہِنْدِ /la:v^{j}-en-hɨnd^{j}/	Ultimate consonant is palatalized. /-en/ is added at the word final position. Suffix /-hɨnd^{j}/ is added at the word final position.
	Fem Pl	لاوِن ہِنْزِ /la:v^{j}-en-hɨnzɨ/	Ultimate consonant is palatalized. /-en/ is added at the word final position. Suffix /-hɨnzi/ is added at the word final position.

Similarly, paradigms for postpositions, and personal pronouns were also built. The number of paradigms developed for the present work is given as follows:
62 paradigms were built for Nouns, 28 paradigms for Verbs, 5 paradigms for Personal Pronouns and 14 paradigms for Adjectives were built. Moreover, a representative list of root words for all these paradigms was prepared. Apart from this, indeclinables like Adverbs, Conjunctions, Vocatives, Interjections etc., were also listed in the root word dictionary.

5. Conclusion

The present work involved building a framework for developing a Morphological Analyzer/Generator for Kashmiri. The first step was categorizing words into parts of speech. This was followed by building paradigms for the grammatical categories on the basis of inflectional properties.
For each grammatical category, different paradigms were created. The main aim was to create as many paradigms within a grammatical category as would be required to include each and every member of that grammatical category, without any exception. The grammatical categories which were observed to be inflected in Kashmiri were Nouns, Pronouns, Verbs, Adjectives and Postpositions, and the work mainly focused on these categories.
The number of paradigms which were observed and built for Nouns was 62, for Verbs 28, for Adjectives 14, and for Pronouns 5. The work also included a representative list of root word dictionary for all the paradigms. Root word dictionaries in simpler terms, are the list of members of each paradigm. Indeclinables (Adverbs, Conjunctions, Interjections, Negatives, etc.) were also included in the morphological dictionary.
The overall premise and the basic advantage of the work is that new or previously unlisted words can easily be included in the root dictionary of the paradigm to which it belongs. The present work is an important step in the direction of creating NLP tools for Kashmiri. The proposed paradigms and framework aim at handling Kashmiri morphological analysis, which in turn can serve as a base for more work in morphological analysis in particular and Kashmiri Natural Language Processing in general.

References

Aarts, J. 1996. A Tribute to W. N. Francis and H. Kucera: Grammatical Annotation. *ICAME Journal*, 20,104–107.
Aho, Alfred V, Ravi Sethi, Jeffrey D. Ullman, and Monica S. Lam.1986. Compilers: Principles, Techniques, and Tools. Pearson.
Antworth, Evan L. 1990. PC-KIMMO: A Two-Level Processor for Morphological Analysis. *Occasional Publications in Academic Computing, Summer Institute of Linguistics, Dallas, Research Center, Palo Alto, CA.* 16.

Aswani, Niraj, and Robert J. Gaizauskas. 2010. Developing Morphological Analysers for South Asian Languages: Experimenting with the Hindi and Gujarati Languages. *LREC*.
Bahl, Lalit R., and Robert L. Mercer. 1976. Part-of-Speech Assignment by a Statistical Decision Algorithm. *International Symposium on Information Theory*, Ronneby: Sweden, 88-89.
Beesley, Kenneth R., and Lauri. Karttunen. 2003. Finite State Morphology: Xerox Tools and Techniques.: Standford. CSLI Publications.
Bharati, Chaitanya V., and R. Sangal. 1995. *Natural Language Processing: A Paninian Perspective*. New Delhi: Prentice-Hall of India.
Bhat, Roopkrishen. 1987. *A Descriptive Study of Kashmiri*. New Delhi: Amar Prakashan.
Crystal, David. *A Dictionary of Linguistics and Phonetics.* Mass: Blackwell-Wiley.
Church, Kenneth Ward. 1988. A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text. *Proceedings of the 2nd Conference on Applied Natural Language Processing*. Austin: University of Texas. 136-143.
Cutting, Doug, et al. 1992. A Practical Part-of-Speech Tagger. *Proceedings of the 3rd Conference on Applied Natural Language Processing*, ACL. 133-140.
DeRose, Steven J. 1988. Grammatical Category Disambiguation by Statistical Optimization. *Computational Linguistics*. 31-39.
Eryiğit, Gülşen, and E. Adalı. 2004. An Affix Stripping Morphological Analyzer for Turkish. *Proceedings of the International Multi- Conference on Artificial Intelligence and Applications*. Innsbruck, Austria.299–304.
Hardie, Andrew. 2004. *The Computational Analysis of Morphosyntactic Categories in Urdu*. Ph.D. Dissertation. University of Lancaster.
Harris, Zellig. Sabbettai. 1962.*String Analysis of Sentence Structure*. The Hague: Mouton.
Heikkila, Juha. 1995. A TWOL-Based Lexicon and Feature System for English. F. Karlsson, A. Voutilainen, J. Heikkila, and A. Anttila. (Eds.) *Constraint Grammar: A Language Independent System for Parsing Unrestricted Text*. Berlin: Mouton de Gruyter. 103 –131.
Hook, Peter E., and O. N. Koul. 1984a. On the Grammar of Derived Transitives and Causatives in Kashmiri. *Aspects of Kashmiri Linguistics*. Ed. O.N. Koul and P. E. Hook. New Delhi: Bahri Publications. 90-122.
Hook, Peter E., and O. N. Koul. 1984b. Pronominal Suffixes and Split Ergativity in Kashmiri. O. N. Koul and P. E. Hook (Eds.) *Aspects of Kashmiri Linguistics*. New Delhi: Bahri Publications. 123-35.
Hook, Peter E., and O. N. Koul. 1992. On the Compound Verb in Kashmiri. *International Journal of Dravidian Linguistics*.1-16.
Huddleston, Rodney. 1984. *Introduction to the Grammar of English*. New York: Cambridge University Press.
Jena, Itisree, et al. 2011. Developing Oriya Morphological Analyzer Using Lt-Toolbox. *Information Systems for Indian Languages.*124-129.
Jurafsky, Daniel, and James H. Martin. 2000. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. London: Prentice Hall.
Kachru, Braj B. 1969. *A Reference Grammar of Kashmiri*. Urbana-Champagne: University of Illinois Press.

❐❐❐