Morphological Analysis of Kashmiri Verbs: A Computational Perspective

Naziya Rasool [*]

Abstract

Morphological analysis is an important part of any natural language processing (NLP) system. Morph analyzer is an algorithm for analyzing the lexical items of a natural language. It can be implemented effectively for the language which is rich in morphology. Morphemes are smallest building blocks that make words in a language.

The present paper discusses the morphological analysis of Kashmiri language. Kashmiri is a highly inflectional language with more than 100 inflected forms for verbs. The word forms of a verb provide information about the different aspects of that particular verb. This information will be useful for developing the morph analyzer of Kashmiri. The verbs in Kashmiri get inflected for tense (present, past, future), number (singular, plural), gender (masculine, feminine) etc. For tense representation verbs in Kashmiri undergo various morphosyntactic changes like vowel changes, addition of suffixes etc. The complexity of word formation makes morphological analysis of Kashmiri verbs a difficult task. The present paper is an attempt at morphological analysis of the Kashmiri verbs. The approach used here for the analysis is Paradigm based approach. A paradigm is a set of inflected forms of a particular lexeme and these word forms are the realizations of that particular lexeme.

Keywords: Morph Analyzer, NLP, Paradigm Based Approach, Inflection.

Introduction

Morphology refers to the mental system involved in word formation or to the branch of linguistics that deals with words, their internal structure, and how they are formed. A major way in which morphologists investigate words, their internal structure and how they are formed is through the identification and study of morphemes, often defined as the smallest linguistic units with a grammatical function.

Morphological analysis is the first phase in any natural language processing application. The output of this phase is used by the syntax- analysis -phase following it. This is also known as the process of providing grammatical information of the word on the basis of properties of the morpheme it contains. Morphological analyzer should accept as input a surface form of a word in a language and return an underlying form divided into morphemes.

Various NLP research groups have developed different algorithms and data structures for morphological analysis and one of the techniques is Paradigm based approach. A paradigm defines all the word forms of a given stem and also provides a feature structure with every word form. The paradigm based approach is efficient for inflectionally rich languages. The table covers a set of roots which means that the roots follow the pattern (or paradigm) implicit in the table for generating their word forms. A Paradigm defines all word forms that can be generated from given stem along with grammatical feature set associated with each word form.

[*] Department of Linguistics, University of Kashmir

Analysis

Kashmiri is morphologically a rich language. Kashmiri is a mix of both the agglutinating and inflection type. It shows both types of morphemic processes across most lexical categories such as Nouns, Pronouns, Verbs, and Adjectives (Wali and Koul, 2009). Verbs in Kashmiri have complex morphology and therefore are very difficult to handle computationally.

Verbs in Kashmiri inflect for number, gender, tense, aspect and mood. The present research analyzes Kashmiri verbs computationally by adopting the paradigm based approach. A paradigm contains all the word forms of a morpheme. Verbs are grouped into different paradigms and analyzed taking various attributes into consideration. All the word forms of a paradigm are converted into computer readable language called as WX notation which in turn makes the analysis easy and systematic. The attributes used here are:

Agr1: Agreement of verb with subject
Agr2: Agreement of verb with direct object
Agr3: Agreement of verb with indirect object
TAM: Tense, Aspect and Mood
Examples:

Paradigm for verb "کھس"

| | | Attributes | | | |
		Agr1	TAM	Agr2	Agr3
کھوُتُس	Kvvyywus	v, m,s,1	v,	any, any, any,	any, any, any
کُھژِس	kQhYjVYQYYs	v, f,s,1	v,	any, any, any,	any, any, any
کھوُتسے	KvvyywsE	v, m,s,1,	v,	any, any, any,	any, s, 2
کُھژسے	kQhYjVYsE	v, f,s,1,	v,	any, any, any,	any, s, 2
کھوُتسو	Kvvyywsv	v, m,s,1,	v,	any, any, any,	any, p, 2
کُھژسو	Kvvyywsv	v, f,s,1	v,	any, any, any,	any, p, 2
کھوُتس	Kvvyywss	v, m,s,1	v,	any, any, any,	any, s, 3
کُھژسس	kQhYjVYss	v, f,s,1	v,	any, any, any,	any, s, 3

کھۆتسکھ	KvvyywsK	v,	m,s,1,	v,	any, any, any,	any, p, 3
کُھڑسکھ	kQhYjVYsK	v,	f,s,1 ,	v,	any, any, any,	any, p, 3
کُھتٮٗ	kQhYwfy	v,	m, p, any,	v,	any ,any, any,	any, any, any
کھِڑ	KjVYQYY	v,	f, p, any,	v,	any, any, any,	any, any, any
کُھتیٚ	kQhYwI	v,	m, p, any,	v,	any, any, any,	any, s, 2
کھڑے	KjVYE	v,	f, p, any	v,	any, any, any,	any, s, 2
کھتٛٗ	Kwiv	v,	m, p, any	v,	any, any, any,	any, p, 2
کھِڑو	KjVYv	v,	f, p, any,	v,	any, any, any,	any, p, 2
کُھتِس	kQhYwis	v,	m, p, any,	v,	any, any, any,	any, s, 3
کھِڑس	KjVYs	v,	f, p, any,	v,	any, any, any,	any, s, 3
کُھتِکھ	kQhYwiK	v,	m, p, any,	v,	any, any, any,	any, p, 3
کھِڑکھ	KjVYK	v,	f, p, any,	v,	any, any, any,	any, p, 3
کھۆتُکھ	KvvyywuK	v,	m, s, 2,	v,	any, any, any,	any, any, any
کُڑِکھ	kQjVYQYYK	v,	f, s, 2,	v,	any, any, any,	any, any, any
کھۆتٮِٚم	KvvyywHm	v,	m, s, 2,	v,	any, any, any,	any, s, 1

كُهژِبم	kQhYjVYHm	v,	f, s, 2,	v,	any, any, any,	any, s, 1
كهۆتِبس	KvvyywHs	v,	m, s, 2,	v,	any, any, any,	a9ny, s, 3
كُهژِبس	kQhYjVYHs	v,	f, s, 2,	v,	any, any, any,	any, s, 3
كهۆتِبكھ	KvvyywHK	v,	m, s, 2,	v,	any, any, any,	any, p, 3
كُهژِبكھ	KvvyywHK	v,	f, s, 2,	v,	any, any, any,	any, p, 3
كُهتِو	kQhYwiv	v,	m, p, 2,	v,	any, any, any,	any, any, any
كهژو	KjVYv	v,	f, p, 2,	v,	any, any, any,	any, any, any
كُهتِوُم	kQhYwivum	v,	m, p, 2,	v,	any, any, any,	any, s, 1
كهژوم	KjVYvm	v,	f, p, 2,	v,	any, any, any,	any, s, 1
كُهتِوُس	kQhYwivus	v,	m, p, 2,	v,	any, any, any,	any, s, 3
كهژوس	KjVYvs	v,	f, p, 2,	v,	any, any, any,	any, s, 3
كُهتِوكھ	kQhYwivK	v,	m, p, 2,	v,	any, any, any,	any, p, 3
كهژوكھ	KjVYvK	v,	f, p, 2,	v,	any, any, any,	any, p, 3
كهۆت	Kvvyyw	v,	m, s, 3,	v,	any, any, any,	any, any, any
كُهژ	kQhYjVY	v,	f, s, 3,	v,	any, any, any,	any, any, any
كهۆتُم	Kvvyywum	v,	m, s, 3,	v,	any, any, any,	any, s, 1

کُہژم	kQhYjVYm	v,	f, s, 3,	v,	any, any, any,	any, s, 1
کُہوَنۡے	kQhYjVYm	v,	m, s, 3,	v,	any, any, any,	any, s, 2
کُہژے	kQhYjVYQYYE	v,	f, s, 3,	v,	any, any, any,	any, s, 2
کھوَنۡو	Kvvyywuv	v,	m, s, 3,	v,	any, any, any,	any, p, 2
کُہژِو	kQhYjVYQYYv	v,	f, s, 3,	v,	any, any, any,	any, p, 2
کھوَنۡس	Kvvyywus	v,	m, s, 3 ,	v,	any, any, any,	any, s, 3
کُہژِس	kQhYjVYQYYs	v,	f, s, 3,	v,	any, any, any,	any, s, 3
کھوَنۡکھ	KvvyywuK	v,	m, s, 3,	v,	any, any, any,	any, p, 3
کُہژِکھ	kQhYjVYQYYK	v,	f, s, 3,	v,	any, any, any,	any, p, 3
کھسان	Ksan	v,	any, any, any,	vA,	any, any, any,	any ,any, any
کھسٕ	KsHQYY	v,	any, s, 1,	iH,	any, any, any,	any, any, any
کھسے	KsE	v,	any, s, 1,	iH,	any, any, any,	any, s, 2
کھسو	Ksvv	v,	any, s, 1,	iH,	any, any, any,	any, p, 2
کھس	Kss	v,	any, s, 1,	iH,	any, any, any,	any, s, 3
کھسٕئس	KsHQs	v,	any, s, 1,	iH,	any, any, any,	any, s, 3

کھسکھ	KsK	v,	any, s, 1,	iH,	any, any, any,	any, p, 3
کھسوِ	KsvQYY	v,	any, p, 1,	iH,	any, any, any,	any, any, any
کھسوے	KsvE	v,	any, p, 1,	iH,	any, any, any,	any, s, 2
کھسوس	Ksvs	v,	any, p, 1,	iH,	any, any, any,	any, s, 3
کھسوکھ	KsvK	v,	any, p, 1,	iH,	any, any, any,	any, p, 3
کھسکھ	KsK	v,	any, s, 2,	iH,	any, any, any,	any, any, any
کھسئَم	KsHQm	v,	any, s, 2,	iH,	any, any, any,	any, s, 1
کھسئُس	KsHQs	v,	any, s, 2,	iH,	any, any, any,	any, s, 3
کھسئَکھ	KsHQK	v,	any, s, 2,	iH,	any, any, any,	any, p, 3
کُھسِو	kQhYsiv	v,	any, p, 2,	iH,	any, any, any,	any, any, any
کُھسِوُم	kQhYsivum	v,	any, p, 2,	iH,	any, any, any,	any, s, 1
کُھسِوُس	kQhYsivus	v,	any, p, 2,	iH,	any, any, any,	any, s, 3
کُھسِوُکھ	kQhYsivuK	v,	any, p, 2,	iH,	any, any, any,	any, p, 3
کھسِ	KsHi	v,	any, s, 3,	iH,	any, any, any,	any, any, any
کھسیم	KsbYlym	v,	any, s, 3,	iH,	any, any, any,	any, s, 1

کهسِس	KsbYlys	v, any, s, 3,	iH,	any, any, any,	any, s, 3
کهسِکه	KsbYlyK	v, any, s, 3,	iH,	any, any, any,	any, p, 3
کهسن	Ksn	v, any, p, 3,	iH,	any, any, any,	any, any, any
کهسنم	Ksnm	v, any, p, 3,	iH,	any, any, any,	any, s, 1
کهسنس	Ksns	v, any, p, 3,	iH,	any, any, any,	any, s, 3
کهسنکه	KsnK	v, any, p, 3,	iH,	any, any, any,	any, p, 3

Conclusion

This paper presented the morphological analysis of Kashmiri verbs using Paradigm based approach. Kashmiri is a morphologically rich language and in case of verbs, number of inflected forms is more in transitive verbs as compared to intransitive verbs. Each inflected form shows change with subject, direct object and indirect object across all forms of tense. This work is a first of its kind and can serve as an input in the direction of developing a morphological analyser for Kashmiri.

References

Bauer, Laurie. 2003. *Introducing Linguistic Morphology,* 2nd ed. Edinburgh University Press.
Booij, Geert. 2005. *The Grammar of Words: An Introduction to Linguistic Morphology.* Oxford UP.
Choudhary, N. K. 2006. *Developing a Computational Framework for the Verb Morphology of Great Andamanese.* Ph.D dissertation. Jawaharlal Nehru University.
Koul, Omkar Nath, and Kashi Wali. 2006. *Modern Kashmiri Grammar.* Springfield: Dunwoody Press.
Koul, Omkar Nath. 2005. *Studies In Kashmiri Linguistics*. Indian Institute of Language Studies.
McCarthy, Andrew C. 2002.*An Introduction to English Morphology: Words and their Structure.* Edinburgh UP.
Prasain, Balaram. 2011. *A Computational Analysis of Nepali Morphology: A Model for Natural Language Processing.* Ph.D Dissertation. Tribhuvan University.
Wali, Kashi, and Omkar Nath Koul. 1997. *Kashmiri: A Cognitive-Descriptive Grammar.* London: Routledge.

❏❏❏