Tense and Agreement in Kashmiri Verbs: A Paradigm Based Approach

Naziya Rasool, Sumaira Nabi and Aadil Amin Kak

Abstract

Natural Language Processing is the process of extracting grammatical information of a word on the basis of morphemes it contains. It is the process of identification, analysis and description of the given language's morphemes and other linguistic units such as root words, stems etc. Morphological analyzer is a tool which identifies and analyzes the internal structure of given word and also extracts the morphological and grammatical information associated with it.

The present work includes the study of tense and agreement of Kashmiri verbs using Paradigm based approach taking only Intransitives into consideration. The verbs in Kashmiri get inflected for tense (present, past, future), number (singular, plural), gender (Masculine, feminine) etc.. This work is the first of its kind and will serve as an important pre-processing tool for developing a machine translation system for Kashmiri.

Keywords: Morphological Analyzer, NLP, Paradigm Based Approach.

Morphology

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. The focus of the present paper is to study the tense and agreement of Intransitive Verbs using Paradigm Based Approach. A paradigm includes all the possible word forms of a root word. Kashmiri verbs show most of the morphological complexities, taking a large number of inflections. It is pertinent to mention here that verbs in Kashmiri inflect for number, gender, person, tense resulting in a number of word forms of a particular stem or root.

Literature Review

So far various morphological analyzers have been developed for different languages using a number of approaches. Some of them are discussed below:

Stochastic Morphological Analyzer using a statistical tagging model has been developed for Japanese in 1994 by Nagata.

Finite State Morphological Analyzer for Hebrew is developed by Yona in 2004.

Morphological Analyzer for Telegu developed by rao.et.al, in 2007 was based on Word and Paradigm method.

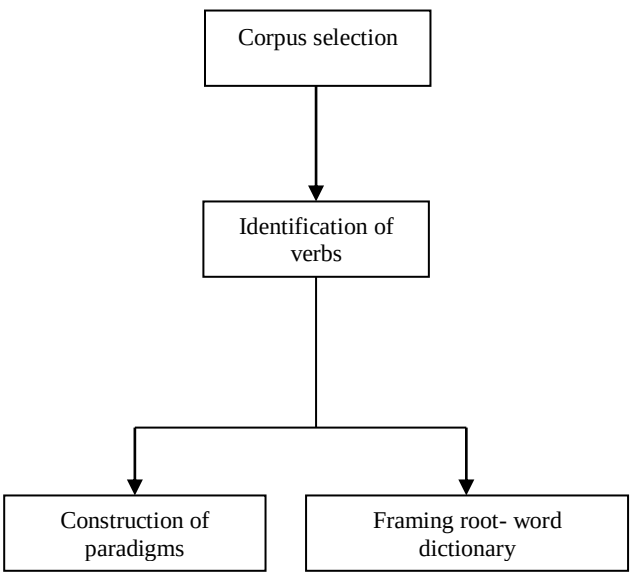
2.4Morphological Analyzer for Marathiwasdeveloped by Bapat, et al., in 2010. Its methodology was based on the use of paradigm-based inflectional system combined with Finite State Machines (FSMs) formodeling the morphotactics.

1. Procedure adopted for the present work

The various steps involved for carrying out the present study are given below:

- Corpus selection from different sources- primary and secondary sources.
- Identification and selection of verbs from the corpus.
- Construction of Paradigms.
- Framing root word Dictionary.

The diagrammatic representation of the procedure adopted for the current work is shown below:



Paradigm Construction

Paradigm construction includes a Paradigm table and grammatical feature set associated with each word form. For the Paradigm construction, imperative form of the verb is taken as root word. The paradigm table includes all the possible word forms of the root word

while grammatical feature set includes Attributes. The analysis of Kashmiri Intransitive Verbs includes construction of Paradigms along with building of Lexicon (list of root words).

Attributes, an important part of the paradigm table includes: (lexcat) lexical category, *TAM* (tense, aspect and mood), *Agr1* (agreement with the subject), Agr2 (agreement with direct object) and Agr3 (agreement with the indirect object). An important observation was that all the three types of agreements showed changes with the morphophonemic changes in the word forms.

For the current work, a number of Kashmiri Intransitive verbs were selected and analyzed. The whole process of paradigm construction and lexicon-building has been shown using the example given in Table 1.This table also includes *WX*-notation (a computer readable Language) of each word form:

Table 1.Paradigm for verb "پک" /pak/ 'walk'

Root word-کھس			Attributes				
Tense	**Word forms**	**Wx-notation**	**Lexcat**	**Agr1**	**TAM**	**Agr2**	**Agr3**
Past	کھۄتُس /kʰotus/	Kvvyywus	v,	m,s,1	v,	any, any, any,	any, any, any
	کھۄڑِس /kʰətsis/	kQhYjVYQYYs	v,	f,s,1	v,	any, any, any,	any, any, any
	کھۄتسے /kʰotsaj/	KvvyywsE	v,	m,s,1,	v,	any, any, any,	any, s, 2
	کھۄڑسے /kʰətssaj/	kQhYjVYsE	v,	f,s,1,	v,	any, any, any,	any, s, 2
	کھۄتسو /kʰotsav/	Kvvyywsv	v,	m,s,1,	v,	any, any, any,	any, p, 2
	کھۄڑسو /kʰətssav/	Kvvyywsv	v,	f,s,1	v,	any, any, any,	any, p, 2
	کھۄتس /kʰotsas/	Kvvyywss	v,	m,s,1	v,	any, any, any,	any, s, 3
	کھۄڑس /kʰətsas/	kQhYjVYss	v,	f,s,1	v,	any, any, any,	any, s, 3
	کھۄتسکھ /kʰotsakʰ/	KvvyywsK	v,	m,s,1,	v,	any, any, any,	any, p, 3
	کھۄڑسکھ /kʰətssakʰ/	kQhYjVYsK	v,	f,s,1 ,	v,	any, any, any,	any, p, 3
	کھۄتی /kʰəti/	kQhYwfy	v,	m, p, 1,	v,	any ,any, any,	any, any, any
	کھۄتی /kʰəti/	kQhYwfy	v,	m, p, 3,	v,	any ,any, any,	any, any, any
	کھۄڑ /kʰatsi/	KjVYQYY	v,	f, p, 1,	v,	any, any, any,	any, any, any
	کھۄڑ /kʰatsi/	KjVYQYY	v,	f, p, 3,	v,	any, any, any,	any, any, any
	کھۄتی /kʰəti:/	kQhYwI	v,	m, p, 1,	v,	any, any, any,	any, s, 2
	کھۄتی	kQhYwI	v,	m, p, 3,	v,	any, any, any,	any, s, 2

	/kəti:/						
	کھڑے /katsaj/	KjVYE	v,	f, p, 1,	v,	any, any, any,	any, s, 2
	کھڑے /katsaj/	KjVYE	v,	f, p, 3,	v,	any, any, any,	any, s, 2
	کھتو /kətiv/	Kwiv	v,	m, p, 1,	v,	any, any, any,	any, p, 2
	کھتو /kətiv/	Kwiv	v,	m, p, 3,	v,	any, any, any,	any, p, 2
	کھڑو /katsav/	KjVYv	v,	f, p, 1,	v,	any, any, any,	any, p, 2
	کھڑو /katsav/	KjVYv	v,	f, p, 3,	v,	any, any, any,	any, p, 2
	گھتس /kətis/	kQhYwis	v,	m, p, 1,	v,	any, any, any,	any, s, 3
	گھتس /kətis/	kQhYwis	v,	m, p, 3,	v,	any, any, any,	any, s, 3
	کھڑس /katsas/	KjVYs	v,	f, p, 1,	v,	any, any, any,	any, s, 3
	کھڑس /katsas/	KjVYs	v,	f, p, 3,	v,	any, any, any,	any, s, 3
	گھتک /kətik/	kQhYwiK	v,	m, p, 1,	v,	any, any, any,	any, p, 3
	گھتک /kətik/	kQhYwiK	v,	m, p, 3,	v,	any, any, any,	any, p, 3
	کھڑک /katsak/	KjVYK	v,	f, p, 1,	v,	any, any, any,	any, p, 3
	کھڑک /katsak/	KjVYK	v,	f, p, 3,	v,	any, any, any,	any, p, 3
	کھوٹک /kotuk/	KvvyywuK	v,	m, s, 2,	v,	any, any, any,	any, any, any
	کڑکھ /kətsik/	kQjVYQYYK	v,	f, s, 2,	v,	any, any, any,	any, any, any
	کھوتھم /kotham/	KvvyywHm	v,	m, s, 2,	v,	any, any, any,	any, s, 1
	گھڑھم /kətsham/	kQhYjVYHm	v,	f, s, 2,	v,	any, any, any,	any, s, 1
	کھوتھس /kothas/	KvvyywHs	v,	m, s, 2,	v,	any, any, any,	any, s, 3
	گھڑھس /kətshas/	kQhYjVYHs	v,	f, s, 2,	v,	any, any, any,	any, s, 3
	کھوتھک /kothak/	KvvyywHK	v,	m, s, 2,	v,	any, any, any,	any, p, 3
	گھڑھک /kətshak/	KvvyywHK	v,	f, s, 2,	v,	any, any, any,	any, p, 3
	گھتو /kətiv/	kQhYwiv	v,	m, p, 2,	v,	any, any, any,	any, any, any
	کھڑو /katsav/	KjVYv	v,	f, p, 2,	v,	any, any, any,	any, any, any
	گھتوم /kətivum/	kQhYwivum	v,	m, p, 2,	v,	any, any, any,	any, s, 1
	کھڑوم /katsvam/	KjVYvm	v,	f, p, 2,	v,	any, any, any,	any, s, 1
	گھتوس /kətivu:s/	kQhYwivus	v,	m, p, 2,	v,	any, any, any,	any, s, 3
	کھڑوس	KjVYvs	v,	f, p, 2,	v,	any, any, any,	any, s, 3

	/ˈkatsoːs/						
	کھٕتِوکھ /kəʰtivakʰ/	kQhYwivK	v,	m, p, 2,	v,	any, any, any,	any, p, 3
	کھٕڑوکھ /katsvakʰ/	KjVYvK	v,	f, p, 2,	v,	any, any, any,	any, p, 3
	کھوٚت /kʰot/	Kvvyyw	v,	m, s, 3,	v,	any, any, any,	any, any, any
	کھٕڑ /kʰəts/	kQhYjVY	v,	f, s, 3,	v,	any, any, any,	any, any, any
	کھوٚتُم /kʰotum/	Kvvyywum	v,	m, s, 3,	v,	any, any, any,	any, s, 1
	کھٕڑم /kʰətsɨm/	kQhYjVYm	v,	f, s, 3,	v,	any, any, any,	any, s, 1
	کھوٚتُج /kʰotuj/	kQhYjVYm	v,	m, s, 3,	v,	any, any, any,	any, s, 2
	کھٕڑِج /kʰətsɨj/	kQhYjVYQYYE	v,	f, s, 3,	v,	any, any, any,	any, s, 2
	کھوٚتُو /kʰotuv/	Kvvyywuv	v,	m, s, 3,	v,	any, any, any,	any, p, 2
	کھٕڑِو /kʰətsɨv/	kQhYjVYQYYv	v,	f, s, 3,	v,	any, any, any,	any, p, 2
	کھوٚتُس /kʰotus/	Kvvyywus	v,	m, s, 3 ,	v,	any, any, any,	any, s, 3
	کھٕڑِس /kʰətsɨs/	kQhYjVYQYYs	v,	f, s, 3,	v,	any, any, any,	any, s, 3
	کھوٚتُک /kʰotuk/	KvvyywuK	v,	m, s, 3,	v,	any, any, any,	any, p, 3
	کھٕڑِک /kʰətsɨk/	kQhYjVYQYYK	v,	f, s, 3,	v,	any, any, any,	any, p, 3
Present	کھسان /kʰasaːn/	Ksan	v,	any, any, any	vA,	any, any, any,	any ,any, any
Future	کھسِ /kʰasɨ/	KsHQYY	v,	any, s, 1,	iH,	any, any, any,	any, any, any
	کھسےج /kʰasaj/	KsE	v,	any, s, 1,	iH,	any, any, any,	any, s, 2
	کھسوٚو /kʰasoːv/	Ksvv	v,	any, s, 1,	iH,	any, any, any,	any, p, 2
	کھسس /kʰasas/	Kss	v,	any, s, 1,	iH,	any, any, any,	any, s, 3
	کھسِس /kʰasɨhas/	KsHQs	v,	any, s, 1,	iH,	any, any, any,	any, s, 3
	کھسکھ /kʰasakʰ/	KsK	v,	any, s, 1,	iH,	any, any, any,	any, p, 3
	کھسوٚو /kʰasvɨ/	KsvQYY	v,	any, p, 1,	iH,	any, any, any,	any, any, any
	کھسوٚے /kʰasoːj/	KsvE	v,	any, p, 1,	iH,	any, any, any,	any, s, 2
	کھسوٚس /kʰasoːs/	Ksvs	v,	any, p, 1,	iH,	any, any, any,	any, s, 3
	کھسوکھ /kʰasoːkʰ/	KsvK	v,	any, p, 1,	iH,	any, any, any,	any, p, 3
	کھسکھ /kʰasak/	KsK	v,	any, s, 2,	iH,	any, any, any,	any, any, any
	کھسِم /kʰashəm/	KsHQm	v,	any, s, 2,	iH,	any, any, any,	any, s, 1

	کھسہٕس /kashəs/	KsHQs	v,	any, s, 2,	iH,	any, any, any,	any, s, 3
	کھسہٕکہٕ /kashək/	KsHQK	v,	any, s, 2,	iH,	any, any, any,	any, p, 3
	کھسِو /kəsiv/	kQhYsiv	v,	any, p, 2,	iH,	any, any, any,	any, any, any
	کھسِوُم /kəsivum/	kQhYsivum	v,	any, p, 2,	iH,	any, any, any,	any, s, 1
	کھسِوُس /kəsivus/	kQhYsivus	v,	any, p, 2,	iH,	any, any, any,	any, s, 3
	کھسِوُکہٕ /kəsivuk/	kQhYsivuK	v,	any, p, 2,	iH,	any, any, any,	any, p, 3
	کھسہٕ /kasi/	KsHi	v,	any, s, 3,	iH,	any, any, any,	any, any, any
	کھسہٕم /kasɨem/	KsbYlym	v,	any, s, 3,	iH,	any, any, any,	any, s, 1
	کھسہٕس /kasɨes/	KsbYlys	v,	any, s, 3,	iH,	any, any, any,	any, s, 3
	کھسہٕکہٕ /kasɨek/	KsbYlyK	v,	any, s, 3,	iH,	any, any, any,	any, p, 3
	کھسن /kasan/	Ksn	v,	any, p, 3,	iH,	any, any, any,	any, any, any
	کھسنم /kasnam/	Ksnm	v,	any, p, 3,	iH,	any, any, any,	any, s, 1
	کھسنس /kasnas/	Ksns	v,	any, p, 3,	iH,	any, any, any,	any, s, 3
	کھسنکہٕ /kasnak/	KsnK	v,	any, p, 3,	iH,	any, any, any,	any, p, 3

The above table shows that each word form undergoes morphophonemic changes and takes different inflections. These inflections vary with number, gender, person and tense.

Consider the first word form in the table **پوٚکُس /pokus/,** the morphophonemic changes reflect that the subject of the word is First Person, Singular and Masculine, and tense is Past. However, the number, gender and person of direct and indirect object can be any (either singular or plural, Masculine/Feminine, first/second/third Person).Similarly for the word form **پُچس /pəʧis/,**Agr1 is First Person, Singular, and Feminine with no change in the tense as well as in Agr2 and Agr3.For the word form پوٚکسے/poksaj/, Agr1 is First Person, Singular and Masculine with no change in Agr2. But Agr3 is Second Person, Singular and gender can be any. The word form پوٚکسو/poksav/, Agr1 is First Person, Singular and Masculine, Agr2 can be any and Agr3 is Second Person, Plural.

The word forms show most of the inflections in the Past tense, with less number of inflections in Future tense and remain same across Present tense. All the morphophonemic changes and their impact on Agr1, Agr2 and Agr3 with respect to Person, number, Gender and Tense is given in the Table 1.

Conclusion

The present paper involved the tense and agreement of Kashmiri Intransitive Verbs using paradigm based approach. The analysis involvedthe paradigm construction and lexicon-building. Paradigms were constructed on the basis of the inflectional properties and morphophonemic changes in the root word (verb). The construction of paradigms for the intransitive verbs led to an important observation that the intransitive verbs in Kashmiri (along with its inflections, etc) havea great deal of information related to subject and object contained in them. In most of the cases the Verb itself could present the sentence.Most of the Kashmiri Intransitive verbs forms taken for the present work behaved idiosyncratically i.e. for each Verb form a separate paradigm had to be constructed.

Future scope

Till date no morphological analyzer has been developed for Kashmiri and this work will act as an important milestone towards this direction which in turn will pave the way for developing a Machine Translation System for Kashmiri.

Bibliography

Bapat, Mugdha, et al. "A paradigm- based Finite State Morphological Analyzer for Marathi." *Proceedings of the Ist workshop on south and Southeast Asian Natural Language Processing (WSSANLP)*,2010.

Bhat, R. (1987). *A descriptive study of Kashmiri*. Amar Prakashan.

Bauer, L. (2003). Introducing linguistic morphology

Baxi, J., Patel, P., & Bhatt, B. (2015). Morphological Analyzer for Gujarati using Paradigm based approach with Knowledge based and Statistical Methods. In *Proceedings of the 12th International Conference on Natural Language Processing* (pp. 178-182).

Goyal, V., &Lehal, G. S. (2008, July). Hindi morphological analyzer and generator. In *First International Conference on Emerging Trends in Engineering and Technology* (pp. 1156-1159). IEEE.

Koul, O. N., &Wali, K. (2004). *Modern Kashmiri Grammar*. Dunwoody Press.

Nagata, Masaki. "A Stochastic Japanese Morphological Analyzer using a forward-DP backward A* N-best search algorithm." *Proceeding of the 15th conference on computational linguistics*, Vol 1.Association for Computational Linguistics, 1994.

Rao G, Uma Maheshwar, et al. "A Telugu Morphological Analyzer." 2007. Centre for Applied Linguistics, University of Hyderabad.

http://sanskrit.uohyd.ac.in/faculty/amba/PUBLICATIONS/papers/ITIC-MA.pdf

Sampson, G. "Alternative Grammatical Coding Systems." *The computational Analysis of English. A corpus-Based Approach*, edited by R.G. Garside. Longman, 1987, pp. 165-183.

Saranya, S. K. (2008). Morphological analyzer for malayalam verbs. *Unpublished M. Tech Thesis, Amrita School of Engineering, Coimbatore*.

Wali, K., Koul, O. N., &Kaula, O. A. (1997). *Kashmiri: A cognitive-descriptive grammar*. Psychology Press.

Yona, Shlomo. *A Finite-State Morphological Analyzer for Hebrew*.2004. U of Haifa. PhD dissertation.