P-ISSN 2708-6453	Pakistan Languages and Humanities Review	April-June 2023, Vol. 7, No. 2
O-ISSN 2708-6461	http://doi.org/10.47205/plhr.2023(7-II)18	[207-216]

Pakistan Languages and Humanities Review
www.plhr.org.pk

RESEARCH PAPER

Functional Morphology: Patterns Rules Sets for Morphological Parsing of Urdu Regular Verbs

Anam Amin[1] Huma Javed[2] Sidra Haroon*[3]

1. M.Phil. Scholar, Department of Linguistics and Communication, University of Management and Technology, Lahore, Pakistan
2. Visiting Lecturer, Department of English Linguistics, University of Okara, Okara, Pakistan
3. Lecturer, Department of Linguistics and Communication, Institute of Liberal Arts, University of Management and Technology, Lahore, Pakistan

***Corresponding Author** Sidra.haroon@umt.edu.pk

ABSTRACT

Affixation of Urdu regular verbs give multiple patterns providing various verbs, varying in meaning and use regarding aspect and gender following similar root verb. Taking into consideration the pattern variations in verbs the aim of this study was to identify morphological patterns of regular Urdu verbs, whereas objective of this research was to explore morphological patterns of stemming in case of Urdu verbs. A corpus of more than 1500 regular verbs is developed. 16 rules sets were developed following 13 different affixation patterns. Out of which 7 specific rules combinations were designed following these combinations in relation to 1 root verb. According to functional morphology these rules set follow a morphological parser for Natural language processing (NLP). The results are significant to design Urdu database of regular verbs corpus for Urdu NLP. This morphological implementation can be used to develop Urdu data base dealing with Functional Morphology and lexicology while dealing Urdu regular Verb forms.

KEYWORDS Affixation, Functional Morphology, Natural Language Processing, Root Verbs, Urdu Regular Verbs

Introduction

This study is based on Functional Morphology of Urdu regular Verbs. Urdu verbs are complex in nature as compared to other grammatical classes. Urdu is considered second most abundant language in regard to its lexical and morphological structures among South Asian Languages (Mehboob and Kortmann, 2004). In Urdu language verbs are represented according to mood, aspect, gender and time. Urdu language verbs show both Causative and In Causative forms when they are defined as infinitive. In Urdu language a 'root' verb usually followed by several affixes or inflections. These root verbs further generate other verbs which some times vary in their meaning and use as well apart of verb form. For example one root word "بن" ، may form different categories of verb from causative to in causative and infinitive to subjunctive even depending aspect and time etc. بنوانا بننا، بنا ، بنے، بنانا،. Moreover, urdu Language attains auxiliaries. The abundance varied auxiliaries/helping verbs represent mood, aspect and time with in a sentence.

Morphology is branch of linguistic that studies the structure of words and their different dictionary forms. A smallest meaningful unit of a word is called morpheme. It could be a complete word ' جگ' or a word element ' نا '. As Urdu language is originated from different languages i.e Arabic, Persian, Turkish, English and Roman. In urdu language so most of the words are made up of these inflections or word elements. These

inflectional patterns are more obvious in Urdu verbs. For example, one infinitive verb **ہونا** is made up of stem **ہو** and it's affix **نا** , this root verb follows different affixation patterns where stem remains same while several verb rules are formed developing new words out of affixation, which is investigated in this study . Functional morphology deals with the morphological system of language independent and dependent part (Haskell, 2004; as cited in Humayon 2006). Present study is set to derive the rules for morphological parsing of Urdu Regular Verbs. A corpus of 1500 Verbs are opted to define the rule sets. Which may help deriving database for FM (Functional Morphology) of Urdu Regular Verbs. According to Humayon (2006), using FM and GF, it is very simple to separate morphology from syntax and semantics to allow a linguistic developer to focus on each component more effectively. (Hardie, 2005)

Morphology is a sub discipline of computational linguistics that studies the internal structure of words. Knowledge of the information that each word conveys inside its structure is a prerequisite for developing various natural language processing applications for a given language. A morphological analyzer attempts to offer a structured representation of a word by breaking it down into the smallest grammatical parts, known as morphemes (Niazi, 2020).

Literature Review

Functional Morphology

FM is a toolkit for developing morphologies in Haskell, a functional programming language created by Markus Forsberg and Aarne Ranta (Forsberg & Ranta, 2004). It is based on the idea of defining morphology using the high expressiveness provided by functional languages. The use of Haskell provides access to powerful programming constructs and a high level of abstraction, which is very useful for capturing natural language generalization. FM library is also a morphological component of Grammatical Framework (GF). The Language Technology Group at Chalmers and Gothenburg University is developing GF, a special-purpose programming language for grammars. GF is a type-theoretic functional programming language. Despite the fact that morphological implementation of a language can be written in GF, but FM gives you more control, freedom, and functionality for defining Haskell's powerful programming constructs enable morphology. FM implementation allowed GF to seamlessly extend it from morphology to syntax and Linguistics. As a result, using FM and GF, it is very simple to separate morphology from syntax and semantics allowing a linguistic developer to focus on each component more effectively (Humayon, 2006).

Syed (2007) discusses Urdu as Morphological rich language (MRL) suggesting MRLs as languages having significant information about syntactic units and their relationships is conveyed at the word level, i.e., word structures are complex and morphological procedures such as inflection and derivation are more common. Although Urdu language processing is very complex because it is influenced by many other languages e.g., Hindi, Persian, Arabic, Sanskrit and English, etc. It's morphemes and constituents tend to be more complex due to derivational inflections (Niazi, 2020). It has been observed that Urdu verbs may contain upto 57 inflected forms (Rizvi and Hussain, 2005; as cited in Niazi 2020).

In Urdu language verbs change to specify gender and stipulate their subjects. Based on the affixations, a distinction is frequently drawn between inflection and derivation when performing morphological analysis. Inflection is the process of adding an affix to a

word without changing its category; for example, "introducing" is an inflected form of "introduce".(R.A. Islam 2011, 2020).

Research methodology

A detailed analysis of more than 1000 regular verbs has been carried out. Following **12** rules are derived to form morphological parsing of these Urdu regular verbs.

1. **نا:Verb + infinitive+ Singular:**

According to this rule the root verb is followed by **affix نا** . these are basic form of verbs. In English infinitive forms are without inflections but in Urdu the infinitive forms are unique because of it's inflection **نا**. As in English infinitive form of base verb see is 'to see', where 'to' is usually categorized as preposition and a separate entity. But in Urdu infinitive form '**دیکھنا**' is one grammatical category. Other examples are **ہونا** derived from **ہو** , **جانا** derived from **جا**.

2. **تا : Verb + Pre+ M+ Singular:**

In Urdu the root verbs are followed by affix **تا** . If we derive continuous form from infinitive verb then **نا** of infinitive is replaced by **تا** to form singular Masculine imperative. It is same in present past and future time. As **ہوتا** is derived from it's root word **ہو** Or infinitive form **ہونا**. Other examples of Singular Masculine imperative forms of Urdu verbs are **دیکھتا** derived from **دیکھ**, **جاتا** derived from root verb **جا**.

3. **تی : Verb + Pre/ Past+ F + Singular:**

In urdu verbs continuous Singular feminine verbs has different affix which is a different pattern from English verb forms. Affix **تی** in urdu defines imperative singular feminine verb in present , past and future times. The root verb is followed by affix **تی** to make Imperative singular female. As **دیکھتی** is derived from root verb **دیکھ**. Other examples are **آتی** from آ. **کھاتی** derived from **کھا** etc. although auxiliaries of present and past defines the time but pattern of verb remains same in both times.

4. **تیں : Verb + pre/past+ F+ Plural:**

تیں affix is when followed by root verb it makes continuous form in plural feminine verb in Present , past and future as **کھاتیں** is derived from **کھا**. The verb pattern is same in present and past form for plural feminine. Other examples are **کرتیں**. Derived from root verb **کر**. **جاتیں** derived from root verb **جا**.

5. **تے : Verb+ Present/ past+ Masculine+ Plural:**

When affix **تے** is proceeded with Urdu verbs it forms plural masculine continuous of past , present . As **کھاتے** is derived from root verb **کھا**. **جاتے** is derived from root verb **جا**. Auxiliary defines time for these verbs in Urdu language. As **کھاتے ہیں**، for present while **کھاتے تھے** . for past.

6. **یا : Verb+ Present / Past+ Masculine+ singular:**

When affix **یا** is followed by regular verb in Urdu it gives a derived form of past. As **کھایا** derived from **کھا** and in present it is followed by auxiliary **ہے** as **کھایا ہے** in assertive form of verb. It has same pattern for both masculine and feminine singular form. Other examples are **پیا** for past and **پیا ہے** for present. **لیا** is used in past and **لیا ہے** in present.

Conclusively this verb pattern is same in present and past time for singular masculine and feminine.

7. ا: Verb + Past+ Masculine+ Plural:

Urdu regular verbs are when followed by affix ا, verbs for past form are derived for both masculine and feminine. For example تلا is derived from a root verb تل by adding an affix ا. In sentences such verbs behave like

لڑکی نے سموسہ تلا۔

امی نے بٹن ٹانکا۔

میں باغ میں ٹہلا۔

8. Verb+ Past/present + Masculine+ Plural/ V+ future + M+ sing ے:

Affix ے is added with regular root verbs to derive plural masculine verbs in both present and past but in present it is usually followed by an auxiliary ہے. As ٹہلے is derived from a root verb ٹہل . This derived verbٹہلے is used in past and future form. In past it will be used only for Plural masculine form as

لڑکے باغ میں ٹہلے۔

But in future time same verb is used for singular masculine following an auxiliary گا.

لڑکا باغ میں پڑھے گا۔

So same verb pattern is used for Past masculine plural and in some situations it is used for future Masculine singular.

9. ی: Verb + Past + Singular + Feminine:

Affix ی is added with regular root verbs to derive plural masculine verbs in both present and past but in present it is usually followed by an auxiliary ہے. As ٹہلی is derived from a root verb ٹہل . This derived verbٹہلی is used in past and future form. In past it will be used only for Plural masculine form as

لڑکی باغ میں ٹہلی۔

10. ئیں : Verb+ Future + Feminine + Plural:

Affix ئیں has different patterns although this affix forms verbs for possessive plural and for feminine plural in future when added to a root verb. As کھیلیں is derived from root verb کھیل. This verb has two different patterns of usage.

لڑکیاں کھیلیں گیں۔

ہم کھیلیں گے۔

ہم کھلیں گیں۔

11. وانا: Verb+ Indirect .Causative+ Singular:

Indirect causative infinitive verbs are derived when affix **وانا** is added to the root verb. As **بنوانا** is derived from root verb **بن** . Other examples are **توڑوانا**

Derived from root verb **توڑ**. **جانچوانا** derived from root verb **جانچ**.

12. انا: Verb+ direct causative + singular:

Direct causative verbs are formed when affix **انا** is added with root verbs. For example **تھامانا** is derived from root verb **تھام**. And **ٹرخانا** is derived from root verb **ٹرخ**, by adding affix **انا**.

13. و : Verb + Imperative+ singular + Masculine/Feminine:

Affix **و** is added when with verb it forms imperative verbs which cause an order or request in present time for singular masculine and feminine. **کھیلو** derived from root verb . **کھیل**

14. ؤ : Verb + Imperative+ singular + Masculine/Feminine:

The root verbs in this rule are followed by affix **ؤ** thus making imperative and imperative forms causing an order or command. For example **کھاؤ** derived from verb **کھا** . and **کھلاؤ**. So in both cases root verb is **کھا**. But in case of **لا۔ کھلا،** is an additional stem following rule pattern **ؤ**. So in this rule**کھلا** can be treated as single root entity to follow morphological rule patterns in functional category. Some other examples are **آؤ، جاؤ، لاؤ، کھاؤ، بناؤ۔**

15. ے: Verb+ Past + Masculine+ Plural / V+ Future+ Masculine+ Singular:

Affix **ے** when added with certain verbs for example **کھلائے** follows causative verbs in Urdu in past aspects. In this case **کھلا** is taken as stem verb instead of it's root verb **کھا**. Such patterns in Urdu language are abundant as in **پلائے، سلائے، لائے، بیٹھائے**. Same verb forms are followed in future singular masculine proceeded by auxiliary **گا**.

16. ئی: Verb + Past + Singular + Feminine:

Verbs following this affixation pattern are usually singular feminine and fall under category of causative forms and many times they are followed by tense auxiliary. i.e **کھلائی، بلائی، پلائی** where root verbs **کھلا، بلا، پلا** are following affixation of **ئی**. In sentence these verbs follow auxiliary **گئی۔**. and follow pattern as **کھلائی گئی، بلائی گئی، پلائی گئی**. But the affix " **ئی** " when comes just after a root verb in case of verb آ "come" forms feminine verb **آئی** in past form. Although these combinations are present either in form of direct affixation with root verbs, or second case is different where root verbs follow double affixation as in case of **کھلائی, بلائی ، پلائی**.

Results and Discussion

Corpus of 1550 regular verbs follow the following set patterns to derive different categories and their verb form. This rule based stemming involves linguistic repertoire of specifically Urdu language which is various in different syntactic patterns. However present study put forth only morphological study for Urdu data base. While functional Morphology deals with it's two parts, 1) Language dependent 2) Language independent, this study deals with language dependent part concerning morphological rules and lexicon to identify various type of system present in Urdu Verb forms.

1. **Following rules: یا،نا، تا، تی، تے، ے/ائے، ی/ائ، تیں،و/اؤ، ئیں/یں :**

The purpose of defining these rules set is to proceed natural language processing for morphological parsing of this regular verbs group. This group of rules st is applicable for root verbs **بلا، رو، بتا، بٹا، بجا، بچا، آزما، اٹھا، بنا، بو، پا، پسوا، پلا، پکا** .

Moreover the presence of diactric "" before suffixation of **یں , و، ی،** forms another verb category, also noticed by Niazi (2020).

Infinitive Verb	Root verb	Verb form	Rules derived
آنا	آ	آنا	نا: V+ Infinitive + Sing
آنا	آ	آتا	تا: V+pre/past+M+sing
آنا	آ	آتی	تی: V+ pre/past +F +sing
آنا	آ	آتے	تے: V+ pre/past +plu +M
آنا	آ	آئے	ے: V+ past + plu +M ے: V+ Future+sing+M
آنا	آ	آئی	ئی: V + past + sing + F
آنا	آ	آتیں	تیں: V + present/ past+ plu + F
آنا	آ	آیا	یا: V + past + M + sing
آنا	آ	آئیں	ئیں: V+ past + plur + F
آنا	آ	آؤ	ؤ:V+impe+sing+M/F

2. **Following Rules with affix patterns نا، تا، تی، تے، ے ، ئی/ای، تیں، یں، ئیں، و .**

This group of verbs shows single character differences in case of **یں** and **ئیں**. These single character differences are also present at phonological level where sense, meaning and syntactic use of particular verb varies.

INFINITIVE	ROOT VERB	VERB FORM	RULE
بدلنا	بدل	بدلنا	نا : V+ Infinitive + Sing
بدلنا	بدل	بدلتا	تا: V+pre/past+M+sing
بدلنا	بدل	بدلتی	تی: V+ pre/past +F +sing
بدلنا	بدل	بدلتے	تے: V+ pre/past +plu +M
بدلنا	بدل	بدلے	ے: V+ past + plu +M ے : V+ Future+sing+M
بدلنا	بدل	بدلی	ی: V + past + sing + F
بدلنا	بدل	بدلتیں	تیں: V + present/ past+ plu + F
بدلنا	بدل	بدلیں	یں: V+ past + plur + F
بدلنا	بدل	بدلئیں	ئیں: V+ past + plur + F
بدلنا	بدل	بدلو	و :V+impe+sing+M+F

3. **Following Affix patterns, و ،ا ،وانا ، ی ، ے ، تیں، ئیں، تے، تی، تا، نا .**

Such type of verbs show stem internal changes in case of causative verbs as is mentioned in table**.** As root verb **خرید** is stemmed to " **خرید +وا** " towards affix **نا**. While root verb+ **ے** identifies two different patterns in syntax. One form is used to represent past without auxiliary for plural Masculine case meanwhile same morphological category is

identified in present syntax following tense auxiliary while second category represent future singular Masculine case following tense auxiliary of future language.

Infinitive verb	Root verb	Verb form	Rule
خریدنا	خرید	خریدنا	نا: V+ Infinitive + Sing
خریدنا	خرید	خریدتا	تا: V+pre/past+M+sing
خریدنا	خرید	خریدتی	تی: V+ pre/past +F +sing
خریدنا	خرید	خریدتے	تے: V+ pre/past +plu +M
خریدنا	خرید	خریدے	ے: V+ past/present + plu +M ے : V+ Future+sing+M
خریدنا	خرید	خریدی	ی: V + past + sing + F
خریدنا	خرید	خریدتیں	تیں: V + present/ past+ plu + F
خریدنا	خرید	خریدوانا	وانا : V + In.Causative+ Sing
خریدنا	خرید	خریدا	ا : V + Past+ M+ Plu
خریدنا	خرید	خریدیں	یں: V+ past/ present/Future + plur + F
خریدنا	خرید	خریدو	و:V+impe+sing+M+F

4. **Following rules with affix: نا،تا، تی، تے، ے ، تیں، ئیں، ی ، وانا، انا، ا، و۔**

This group of Urdu Regular verbs distinguishes because of presence of causative forms. There are two causative forms direct causative and indirect causative. In this group suffix **وانا** and **انا** makes singular in.causative and causative forms respectively in Urdu regular verbs for singular categories. Although many plural categories may present for certain regular verb forms but this group is limited to singular cases to avoid irregular distinguishes among verb categories.

Infinitive form	Root verb	Verb form	Rules
پڑھنا	پڑھ	پڑھنا	نا: V+ Infinitive + Sing
پڑھنا	پڑھ	پڑھتا	تا: V+pre/past+M+sing
پڑھنا	پڑھ	پڑھتی	تی: V+ pre/past +F +sing
پڑھنا	پڑھ	پڑھتے	تے: V+ pre/past +plu +M
پڑھنا	پڑھ	پڑھے	ے: V+ past /present+ plu +M/ ے : V+ Future+sing+M
پڑھنا	پڑھ	پڑھی	ی: V + past + sing + F
پڑھنا	پڑھ	پڑھتیں	تیں: V + present/ past+ plu + F
پڑھنا	پڑھ	پڑھوانا	وانا : V + In.Causative+ Sing
پڑھنا	پڑھ	پڑھانا	انا:V+ direct causative + sing
پڑھنا	پڑھ	پڑھا	ا : V + Past+ M+ Plu
پڑھنا	پڑھ	پڑھیں	یں : V+ present/Future + plur + F
پڑھنا	پڑھ	پڑھو	و:V+impe+sing+M+F

5. **Following Patterns with نا، تا، تی، تے، ےاے، تیں، ئیں، یائی، ا، و :**

Infinitive form	Root verb	Verb form	Rules
پھسلنا	پھسل	پھسلنا	نا: V+ Infinitive + Sing
پھسلنا	پھسل	پھسلتا	تا: V+pre/past+M+sing
پھسلنا	پھسل	پھسلتی	تی: V+ pre/past +F +sing
پھسلنا	پھسل	پھسلتے	تے: V+ pre/past +plu +M
پھسلنا	پھسل	پھسلے	ے: V+ past + plu +M/ ے : V+ Future+sing+M

پھسلنا	پھسل	پھسلی	ی: V + past + sing + F
پھسلنا	پھسل	پھسلتیں	تیں: V + present/ past+ plu + F
پھسلنا	پھسل	پھسلا	ا : V + Past+ M+ Plu
پھسلنا	پھسل	پھسلیں	یں: V+present/ past/ Future+ plur + F
پھسلنا	پھسل	پھسلو	و:V+impe+sing+M+F

6. Rules following Affix Patterns: نا،تا، تی، تے، ے ، تیں، ئیں، ی، ا، انا، و :

This group of regular root verbs follows only direct causative forms while all imperative conditions are present in this class of verbs.

Infinitive verb	**Root Verb**	**Verb form**	**Rules**
جاگنا	جاگ	جاگنا	نا: V+ Infinitive + Sing
جاگنا	جاگ	جاگتا	تا: V+pre/past+M+sing
جاگنا	جاگ	جاگتی	تی: V+ pre/past +F +sing
جاگنا	جاگ	جاگتے	تے: V+ pre/past +plu +M
جاگنا	جاگ	جاگے	ے: V+ past/present + plu +M
			ے : V+ Future+sing+M
جاگنا	جاگ	جاگی	ی: V + past + sing + F
جاگنا	جاگ	جاگتیں	تیں: V + present/past+ plu + F
جاگنا	جاگ	جاگانا	انا : V + direct. causative+ Sing
جاگنا	جاگ	جاگا	ا : V + Past+ M+ Plu
جاگنا	جاگ	جاگیں	یں: V+ past + plur + F
جاگنا	جاگ	جاگو	و:V+impe+sing+M+F

The rules defined above are applied on 28K Urdu kids' corpus and 38K Urdu corpus based on Udru academic and fiction readings. Following formula was applied to evaluate the probability ratio of relevant verbs into the corpus.

$$P(A) = \frac{\text{Number of favourable Outcome}}{\text{Total number of Outcomes}}$$

The p-value of all above mentioned six rules for Urdu verb affixations and their categories is found less than 0.05 in collective Urdu corpus of 66K corpus. However, the significant value of infinitive causative forms are found 0.08 in the corpus of 66K. Which is asuumed that the occurrence of in.causative forms is found less in Urdu corpus. However, the above 15 derived rules are found in 1550 Urdu regular verbs. These verbs were found in the corpus of Urdu 66K word.

Conclusion

Urdu language has abundance of concatenative inflectional morphological system. This study is helpful to apply Urdu morphology and syntax data base in Urdu Natural Language Processing (NLP) and its morphological parsing of Urdu regular Verbs. Present study has presented the variation of Urdu regular verbs and has drawn six concrete rules drawn out of 16 types of morphological parsing. Future research draws its implications to develop syntax based Urdu morphological parsing in case on Urdu regular verbs.

References

Afzal, M., & Hussain, S. (2001.). Urdu computing standards: Development of Urdu Zabta Takhti (UZT) 1.01. *Proceedings. IEEE International Multi Topic Conference, 2001. IEEE INMIC 2001. Technology for the 21st Century*.

Forsberg, Markus & Ranta, Aarne. (2004). Functional morphology. *Sigplan Notices - SIGPLAN*. 39. 213-223.

Hardie, A. (2005). Automated part-of-speech analysis of Urdu: conceptual and technical issues. In Y. Yadava, G. Bhattarai, R. R. Lohani, B. Prasain, & K. Parajuli (Eds.), *Contemporary* issues in Nepalese linguistics (pp. 49-72). Linguistic Society of Nepal

Humayoun, M., Hammarström, H., & Ranta, A. (2022). Urdu Morphology, Orthography and Lexicon Extraction. *ArXiv, abs/2204.03071*.

Islam, R. A. (2011). *The Morphology of Loanwords in Urdu: the Persian, Arabic and English Strands.* PhD dissertation submitted to Newcastle University.

Khan, T. A. (2020). Morphological Integration of Urdu Loan Words in Pakistani English. *English Language Teaching*, *13*(5), 49-53

Mehboob, A. & Kortmann, B. (2004). *Pakistani English: Morphology and Syntax. A Handbook of varieties of English: A multimedia reference tool*. New York: Mouton de Gruyter.https://doi.org/10.1515/9783110175325.2.1045

Niazi, A. (2018). Morphological Analysis of Urdu Verbs. 10.1007/978-3-319-75477-2_19.

Rahman,T. (2004), *Language Policy and Localization in Pakistan: Proposal for a Paradigmatic Shift, Crossing the Digital Divide*, SCALLA

Rizvi, S., & Hussain, M. (2005). Analysis, Design and Implementation of Urdu Morphological Analyzer. *2005 Student Conference on Engineering Sciences and Technology*, 1-7.

Sara, H., & Sarmad, H. (2004). *Finite-State Morphological analyzer for Urdu.* National University of Computer & Emerging Sciences.

Syed, A. Z. (2013). *Redefining Urdu Morphology and Grammar for the Development of an Integrated Sentiment Analysis Framework*. Department of Computer Science and Engineering University of Engineering and Technology, Lahore, Pakistan