

SOCIAL SCIENCE AND HUMANITIES

Manuscript info:

Received September 16, 2018., Accepted October 18, 2018., Published October 30, 2018.

MORPHOLOGICAL MARKUP ANG LINGUISTIC MODEL

Khamrayeva Shakhlo Mirdjanovna,
Teacher of Bukhara State University,
Bukhara city, Uzbekistan.
hamroyeva81@mail.ru



<http://dx.doi.org/10.26739/2573-5616-2018-10-2-19>

Abstract: In this article describes the linguistic basis of the morphological markup of the Uzbek language corpus, the way to select the wordform tags and the creation of linguistic models for tags. For working at the tendency of morphological markup of texts in Uzbek language A.Kahhar's story called "Patient" does the object of basic. There are wordform, lemma and tag in the system of morphological markup and character of them in Uzbek in Uzbek language and process of lemming flective and agglutinative language are explained. For morphological markup of every independent part of speech necessary tags and making linguistic models of it are basic task of work. For making system of tags of morphological markup, to compose paradigm of characteristic grammatical meaning and words, to choose the most suitable variation from different theories are basic purpose of work. There is need for emphasizing about being given lemma in sign <***>, tags in sign [***]. There is separate place for affixation which is in structure of lemma and is not in structure of lemma in article.

Key words: word, wordform, lemma, the morphological markup, tag, linguistic model

Recommended citation: Khamrayeva Shakhlo. MORPHOLOGICAL MARKUP ANG LINGUISTIC MODEL. 9-10 European Journal of Research P. 187-198 (2018).

Introduction. In Uzbek language, working at specific tendency of morphological markup of the texts does the object of manual for markup of story called "Patient". There are wordform, lemma and tag in the morphological markup system. The wordform is the morphological identity in the selected text. The first level of markup of the wordform is lemming, it means that to indicate lexical form of the wordform. In the markup of the flective languages, the most difficult level is to lemma, it means that the lexical form of word is combined with wordform as a tag. Because in the flective languages, grammatical meaning of wordform is mingled with stem of the word.

In agglutinative language, lemming is rather easy than fleective ones. The part of wordform without grammatical form equals to stem or basic lemma. In markup, lemma is given in sign <*>. If in all of the parts of speech lemming is like that, it means that “the part of word stem-basis equals to lemma”, in verb, verb-lemma is expressed in second personal and order conditional of verb. In dictionary articles, verb is given in form of the name of movement: <to go>. This form is not suitable for corpus, because in the text of corpus the version of <go> is searched not <to go>. It is true that to give the lemma form of verb-phrase with “to” supplementary, because there is no meaning to lemma the phrase <tire out> which was <was tired out>. When user searches phrase from dictionary, he/she uses from semantic layout, not morphological marking. That is why it is rather comfortable to markup the phrase “to tire out” than “was tired out”. The other parts of speech are written in sign <*> as a form of stem-basis and lexeme: **in the morning** <morning>. Of course, there is point of simplifying stem in parts of speech. So, in simplifying process, the stem is altered to lemma without changing. Dictionary form which is in noun is not introduced to lemma, because this form does not influence to meaning and in layout little girl is reflected in form <girl>. Some forms are apart from it. For example, kinder garden, rug, coverlet equals to lemma without changing. In the process of markup, from 5 to 10, sometimes more than that morphological tags are required for every wordform. Linguistic modeling tags are expedient, because in linguistic model morphological tag is indicated in abbreviation form. Forms of special linguistic model are done for every part of speech markup. In the tag – linguistic models of morphological markup and supplementary of work, you can introduce in certain order.

Tags of verb and lemming it. In working at linguistic model of markup of the verb, following grammatical categories are accounted and necessary tags are written according to verb-lemmas and wordform. The name of parts of speech: verb = [v]. According to lexical and grammatical peculiarity, followings are separated:

- a) independent verb = [independent v.];
- b) dependent verb = [dependent v.]:
 - assistant verb = [assistant v.];
 - auxiliary verb = [auxiliary v.];
 - incomplete verb = [incomplete v.];

As a basic vocabulary meaning groups of verb:

1. verb of movement = [movement v.];
2. verb of speech = [speech v.];
3. verb of state = [state v.];
4. verb of effectual activity = [effectual activity v.];
5. verb of thought = [thought v.];
6. verb of interaction = [interaction v.] are separated to

lexical and semantic groups.

According to attitude to object of movement, verb is separated to transitive and intransitive; according real or irreal of verb is separated to positive and negative.

Linguistic model of category of voice has following shape:

1. exact voice = [exact vo.]
2. self voice = [self vo.]
3. passive voice = [passive vo.]
4. increasing voice = [increasing vo.]
5. togetherness voice = [togetherness vo.]

According to this structure of verb:

1. ordinary verb = [ordinary v.];
2. composite verb = [composite v.];
3. couple verb = [couple v.];
4. repeated verb = [repeated v.].

Derivation of verb is separated to derived verb and basic verb as a tag. Affixation and composition method of forming is not written as a tag, because the information about structure of verb is enough: **қонлади** <қонла> [ф], [сод. ф.], [я.ф.] (**slaughtered** <slaughter> [v.], [basic v.], [derived v.]. In this tag there is some information about ordinary derivation of verb formation “қонламоқ” (“to slaughter”), basic verb comes in to existence only by way of affixation. So, there is no necessity to emphasize the structure of compositional verb: **савдо қиладиган** <савдо қил> [ф], [қўш. ф.], [я.ф.] (**sold** <sell> [v], [composite v.], [derived verb]).

It's right that, there is no need for explanation about the structure of verb. The category of verb movement is written as a tag, parts of basic and auxiliary verb are separated. **оғриб қолди** <оғриб қол> [ф], [муст. ф.], [холоат ф.], [ў-сиз ф.], [б-ли ф.], [а. н.], [сод. ф.], [т.ф.], [х.м.], [ў.з.], [III ш.б.], [хар. т.], <оғриб> [ет. ф.], <қолди> [қўм. ф.] (**have ached** <ache> [v], [independent. v.], [state v.], [intransitive. v.], [positive v.], [active. v.], [ordinary. v.], [basic v.], [exact v.], [past t.], [III person c.], [хар. т.], <have > [auxiliary. v.], <deviated > [assistant. v.]).

As a tag there are category of movement: basic verb and auxiliary verb forms in the text. Wordform of the category of verb is lemmed in shape of verb-lemma. The tag of the category of verb of interaction has following model:

- 1) category of case:
 - category of notification = [c. notification];
 - category of condition = [c. condition];
 - category of order = [c. order];
 - category of aim = [c. aim].
- 2) tense category:
 - past tense = [past t.];
 - present tense = [present t.];
 - future tense = [future t.].
- 3) quantity category:
 - first person singular/plural = [I p s/p];
 - second person singular/plural = [II p s/p];
 - third person singular/plural = [III p s/p].

In the text, the phrase which expressed meaning of movement-state was named as a verb-phrase [v. phrase] and tag of verb-phrase was combined to that identity.

Marking out particularities of noun. At the first stage to lemma and mark out the noun, it's general grammatical meaning is major. Proper or *homogeneous* noun and their types are defined according to the explanation of the only kind of the certain group or general name of this kind of such subjects. This grammatical point is designed as tag below:

The name of parts of speech: noun = [noun]. Lemming and marking of private noun's words base on primary common grammatical meaning. One species of the substance or the common denominator of these species has identified the spiritual groups of the most prominent and related species and their type is determined. This grammar is labeled as a tag. Name of part of sentences: Noun = [n].

I. Meaningful groups of Nominal noun [nominal noun] = [n.n].

1. [Personal name] = [P.n];
2. [Geographical name] = [G.n];
3. [Name of organization, founding, office] = [OFO.n];
4. [Heavenly lamps name] = [HL.n];

1. [History date or holiday's name] = [HH.n];
2. [Animals name] = [A.n];
3. [Product name] = [Pr. N];
4. [Divine thoughts name] = [D.n];

Marking form: **Абдуганибой** <Абдуганибой> [от], [ат. от.], [шахс н.], [б. к.], [бирл. с.] (**Abduganibay** <**Abduganibay**> [noun], [P. name], [p. noun], [singular n.]

II. Meaningful groups of Similar noun: [similar noun] = [s.n]

1. [abstract noun] = [a.n];
2. [exact noun] = [e.n].

Exact noun is separated to VMG(1):

1. [Substances name] = [sub.n]
2. [Personal name] = [per.n]
3. [Describe the person from relative side] = [rel.]
4. [Describe the person from age side] = [age]
5. [Describe the person from profession side] = [prof.]
6. [Describe the person from gender side] = [gen.]
7. [Describe the person from residence side] = [res.]
8. [Describe the person from position side] = [pos.]
9. [Describe the person from social position side] = soc.pos.]
10. [object's name] = [o.n]
11. [Plant's name] = [pl.n]
12. [Place's name] = [place n]
13. [Quantity's name] = [q.n]
14. [Name of organization and founding] = [OF.n]
15. [Time's name] = [t. n]
16. [Activity's name] = [act. n]

Marking form: **паркда** <**парк**> [от], [тур. от], [ан.о.], [ў.ж.н.], [я.о.], [с.о.], [т.о.], [ў.п.к.], [бирл. с.] (**in the park** <**парк**> [noun], [homogeneous n.], [exact n.], [OF. n.], [derived n.], [sin. n.], [s.n.], [ў.п.к.], [singular n.]

According to the unitary form of a subject or a substance of a subject:

1. [single noun] = [sin. n];
2. [collective noun] = [col. n].

Marking form: **омборда** <**омбор**> [от], [тур.от], [ан.о.], [ў.ж.н.], [я.о.], [с.о.], [т.о.], [ў.п.к.], [бирл. с.] (**in warehouse** <**warehouse**> [n], [homogeneous n.], [e. n.], [place n.], [sin. n.], [s.n.], underived n.], [singular n.]

According to the structure of the noun it divides into simple, mixed, pair, repetition types are separated:

1. [simple noun] = [s.n];
2. [mixed noun] = [m.n];
3. [pair noun] = [pa.n];
4. [repetition noun] = [r.n].

Marking form: **гўдакнинг** <гўдак> [от], [тур.от], [ан.о.], [ш.н.], [ёш ЛМГ], [я.о.], [с.о.], [т.о.], [қар.к.], [бирл.с.] (**baby's** <**baby**> [n], [homogeneous n.], [per. n.], [age VMG], [sin. n.], [s.n.], underived n.), [singular n.]).

The noun is subdivided into sub and legislative groups:

1. [sub noun] = [sub n];
2. [legislative noun] = [l.n].

Marking form: **ғашлиги** <ғашлик> [от], [тур.от], [мавх.о.], [я.о.], [с.о.], [**яс.о.**], [б.к.], [бирл.с.], [III ш.б.] (**disgust** <**disgust**> [**n.**], [homogeneous n.], [a. n.], [s.n.], [derived n.], [singular n.]).

In describing case of noun, the basic form of ownership and the category in describing the word form is their grammatical form as follow:

I. Model of the category of case:

1. [nominative case] = [n.c];
2. [genitive case] = [g.c];
3. [accusative case] = [ac.c];
4. [dative case] = [d.c];
5. [dative case] = [d.c];
6. [ablative case] = [ab.c];
7. [prepositional case] = [pr.c].

II. Category of numerals:

1. [singular] = [s];
2. [plural] = [p].

III. Category of ownership:

1. [first person] = [I p.];
2. [second person] = [II p.];
3. [third person] = [III p.].

As you can see, lemming and marking of pronouns each pronoun in the form of the word should annex at least 6 (tegs) explanations.

Marking form: **отасининг** <ота> [от], [тур.от], [ан.о.], [ш.н.], [қар. ЛМГ], [я.о.], [с.о.], [т.о.], [**қар.к.**], [III ш.], [бирл.с.] (**father's** <father> [noun], [homogeneous n.], [e.n.], [per.n.], [gen. VMG], [derived n.], [sin.n.], [s.n.], [**қар.к.**], [III person c.], [singular n.]

In the marking of pronouns the problem of morphological ammonia is upright. “it” is morphological ammonia as personal and demonstrative pronouns. Only in the context of this pearl is the meaning. Although the world’s most popular corpus system has been used for automated demolition, the problem of morphological well-being is not solved without human intervention.

It is possible to say that the lexical and morphological well-being is marked and one of the marks will be written as demonstrative and personal pronouns to solve the problem of ammonia. This kind of approach is just morphological ammonia. The problem of lexical tranquility will be discussed later in the next case if necessary.

During the marking the shape of cases there is a problem in the corps: the matter is unmarked using of cases' shapes. Unmarked used case is thought as nominative case by corps.

In the text as a lexical forms of the noun, the following meanings of linguistic forms are assimilated as labeling because of the diminishing of the forms of personal relationship:

1. [spoil meaning] = [sp.m];
2. [humiliate meaning] = [hum.m];
3. [diminish meaning] = [dim.m];
4. [sorrow meaning] = [sor.m.]

Marking forms: **саватчаларини** <сават> [от], [тур.от], [ан.о.], [б.н.], [я.о.], [с.о.], [т.о.], [т.к.], [III ш.], [кўп.с.], [**кич. ЛIII**] (**baskets** <**basket**> [noun], [homogeneous n.], [e.n.], [sub.n.], [derived n.], [sin.n.], [s.n.], [т.к.], [III person c.], [plural n.], [**dim.m.**])

The transposition of the other parts of speech is also considered in regulation. In the text is seen off transposition of adjective to noun. Tag form as following: [transposition adjective-noun] = [transpos. adj.-n.].

Marking form: касални <касал> [Транспоз: сиф→о], [тур.от], [ан.о.], [ш. н.], [я.о.], [с.о.], [т.о.], [туш. к.], [бирл. с.]. (**ill** <**ill**> [Transpose: adj→n], [homogeneous n.], [e.n.], [per.n.], [derived n.], [sin.n.], [s.n.], [туш. к.], [singular n.]).

General issues of the adjective.

In article, independent forms of speech: for verb, noun, adjective, pronoun, adverb tag – problem of making model has been solved. To lemma word in adjective part of speech is done by separating <*> sign of stem-basic from wordform and linguistic model of it is signed in form “adjective” = [adj.]: far <far> [adj.].

Markup of adjective-wordform will begin by signing general grammatical meaning as a noun part of speech. In literatures, according to express peculiarity of sign of adjective is divided into two groups (1):

- 1) Entirely adjective = [ent.a.]
- 2) Relatively adjective = [rel.a.]

Form of markup: **йирок** <йирок> [сиф.], [ac. c.], [о.д.], [с.с.], [т.с.], [хyc. ЛМГ] (**far** <far> [adj.], [ent.a.], [pos.d.], [b.adj.], [pec. VMG])

In the textbooks and manuals VMG of adjective are separated to 6, 8 and sometimes 9 in the markup of the text. VMGs of adjective which are in the description are based to definition which is totally encompassed (1), because in morphological markup, to show exact meaning bases on perfection of semantic markup. VMGs adjective are demonstrated in this form as a tag:

1. Adjective of peculiarity = [pec. VMG]
2. Adjective of state = [st. VMG]
3. Smell denoting quality = [SDQ];
4. Color scheme denoting quality = [CSDQ];
 - adjective of natural state = [nat. s. VMG];
 - adjective of physical state = [phy. s. VMG];
 - adjective of mental state = [men. s. VMG];
 - adjective of financial state = [fin. s. VMG];
 - adjective of temperature state = [tem. s. VMG];
 - adjective of clearness state = [cl. s. VMG];
 - adjective of new-oldness state = [n-old. s. VMG];
 - adjective of wet-dryness state = [w-dry. s. VMG];
 - adjective of quiet-unquiet state = [q-unq. s. VMG];
 - adjective of mature state = [mat. s. VMG];
5. Taste denoting quality = [TDQ];
6. Shape denoting quality = [SHDQ];
7. Measure denoting quality = [MDQ];
8. Position denoting quality = [PDQ];
9. Time denoting quality = [TiDQ]

Marking form: **йирок** <йирок> [сиф.], [ac. c.], [о.д.], [с.с.], [т.с.], [хyc. ЛМГ] (**far** <far> [adj.], [ent.a.], [posd.d.], [b.adj.], [pec. VMG])

The quality degree of category labels:

- 1) positive degree = [pos.d.];
- 2) comparative degree = [comp.d.];
- 3) superlative degree = [sup.d.].

Marking form: **каттик** <каттик> [сиф.], [ас. с.], [од.д.], [с.с.], [т.с.], [хус. ЛМГ]; (hard <hard> [adj.], [ent.a.], [**pos.d.**], [b. adj.], [underived adj.], [pec. **VMG**]).

The structure and derivation of quality are also examined by the scale. Composition type:

- 1) bare adjectives = [b. adj.];
- 2) complex adjectives = [c. adj.];
- 3) pair adjectives = [p. adj.];
- 4) rolling adjectives = [r. adj.] form, quality construction

divides:

- 1) core quality = [c.q.];
- 2) legislative quality = [l.q.].

Marking form: **йўгон** <йўгон> [сиф.], [ас. с.], [од.д.], [**с.с.**], [**т.с.**], [ш. ЛМГ]; (**quiet** <quiet> [adj.], [ent.a.], [pos.d.], [**b. adj.**], [**derived adj.**], [pec **VMG**]).

The quality category has a vowel that does not have a wide range of syntactic categories. Qualitative category of ownership “link the quality to the preceding word and refer to the meaning of nomenclature” in this style privatized. The category of predicate indicates its capability in quality in limited degree. In many cases, the predicates category in the quality range will be released.

Forming adjective to noun marked as adjacent quality = [adj.q].

In story text *Сотиболди қизчасини ўлик ёнидан олиб бошиқа ёққа ётқизаётганда қизча уйғонди* (when Sotiboldi took his daughter from side of dead to another place, little girl waked up.) in this sentences word “o`lik” (dead) is adjacent quality, because the cartridge agreement has been signed without any indication. The sign of word-formation finds reflection in this marking.

Marking form: **ўлик** <ўлик> [сиф.], [нис.с.], [од.д.], [яс.с.], [с.с.], [жис.х. ЛМГ], [**отл.с.**] (**dead** <dead> [adj.], [rel.adj.], [pos.d.], [derived adj.], [b.adj.], [phy.s.VMG]).

The category of predicates refers to “[predic.]” tag: **тузукман** <тузук> [сиф.], [ас. с.], [о.д.], [с.с.], [т.с.], [хус. ЛМГ], [**кес.**] (**I`m healthy** <healthy> [adj.], [ent.a.], [pos.d.], [b.adj.], [underived adj.], [pec. **VMG**], [**pred.**])

Comments on marking pronouns. In markup of pronoun part of speech, vocabulary meaning groups, composition, form, category of ownership, category of numerals are combined as a tag. Tag form of these grammatical demonstrations, linguistic models have following forms: Pronoun part of speech = [pr.].

1. The group by meaning:

- 1) personal pronoun = [per. pr.]
- 2) demonstrative pronoun = [dem. pr.]
- 3) reflexive pronoun = [ref. pr.]
- 4) suspicion pronoun = [sus.pr.]
- 5) interrogative pronoun = [inter. pr]
- 6) contemplate pronoun = [cont. pr]
- 7) negative pronoun = [neg. pr.].

2. Types according to adjustment:

- 1) simple pronoun = [sim. pr];
- 2) compositive pronoun = [com. pr];
- 3) pair pronoun = [pair pr];
- 4) repetitive pronoun = [rep. pr].

3. Formation:

- 1) basic pronoun = [bas. pr.]
- 2) derived pronoun = [der. pr.]

4. Category of pronouns case:

1. [nominative case] = [n.c];
2. [genitive case] = [g.c];
3. [accusative case] = [ac.c];
4. [dative case] = [d.c];
5. [ablative case] = [ab.c];
6. [prepositional case] = [pr.c];

5. Category of numerals:

1. [singular] = [s];
2. [plural] = [p].

6. Category of ownership:

1. [first person] = [I p.];
2. [second person] = [II p.];
3. [third person] = [III p.].

As you can see, lemming and marking of pronouns each pronoun in the form of the word should annex at least 6 (tags) explanations. Marking form: **my** <**my**> [о.л.м.], [кўр.о.л.м.], [с.о.л.м.], [туб о.л.м.], [б.к.], [бирл. с.] (<**this**> <**this**> [pr.], [dem. pr.], [sim. pr.], [underived.], [s.]).

In the marking of pronouns the problem of morphological ammonia is upright. “it” is morphological ammonia as personal and demonstrative pronouns. Only in the context of this pearl is the meaning. Although the world’s most popular corpus system has been used for automated demolition, the problem of morphological well-being is not solved without human intervention. It is possible to say that the lexical and morphological well-being is marked and one of the marks will be written as demonstrative and personal pronouns to solve the problem of ammonia. This kind of approach is just morphological ammonia. The problem of lexical tranquility will be discussed later in the next case if necessary.

Marking out of adverbs: problem and solution. Adverb is the part of speech which is caused a lot of discussion and agreements among linguistics. Even though this part of speech was seriously observed by European, Russian and Uzbek scientist, the sameness of explanation of this part in the textbooks is not achieved yet. Although, the latest idea of Asadov about adverb in our linguism is considered essential.

While marking-out this part of speech of the mark-out called “Patient”. In this observation 1175 words that are suspected to be included in adverb is observed, words text are marked-out to be whether adverb or not based on this timetable. Vocabulary meaning group, formation and deviation of adverb categories are taken as tag.

Adverb is included to word form in type below:

- 1) The name of part of speech: [adverb] = [adv.].
2. Vocabulary meaning groups of adverb are defined:
 - 1) time;
 - 2) foundation;
 - 3) state;
 - 4) quantity;
 - 5) aim;
 - 6) cause.
1. The types according to adverb formation:
 - 1) simple adverb = [sim. adv.];
 - 2) compositive adverb = [comp. adv.];
 - 3) pair adverb = [pair adv.];

4) repetitive adverb =[rep. adv.].

4. Action of adverb:

1) basic adverb = [basic adv.]

2) derived adverb = [derived adv.]

The marking-out appearance of adverbs in the text:

хозир <хозир> [рав.], [п. рав.], [с. рав.], [туб рав.] (now <now> [adv.], [simple adverb], [exact adverb]).

Conclusion. It can be noted that the system of morphological markup is composed of a unit as a word form, a lemma, and tags.

References

1. Asadov T.H. Prichastiye v sisteme chasteq rechi. (Participle in the system of parts of speech) Dissertatsiya na soiskaniye uchenoy stepeni kandidata filologichaskix nauk. Tashkent - 2009.
2. Sayfullaeva R. and others. Hozirgi o`zbek adabiy tili. (Modern Uzbek language). Uchebnoye posobiye. - Tashketn, "Nauka i texnologiya", 2009. - 416 s.
3. Zakharov V.P., Bogdanov S.Yu. Корпусная лингвистика. (Corpus Linguistics). Irkutsk, IGLU, 2011.
4. Zakharov V.P. Korpusniye issledovaniya. Leksiya 3. Corpus studies (Elektronniy resurs) - docplayer.ru/64134811-Lekciya-3-korpusnye-issledovaniya.html