Appendix 4 : The tagging guidelines

The guidelines given in this appendix are identical to the final set of tagging guidelines that were devised during the phase of manual tagging discussed in Chapter 4, with two exceptions. Firstly, they have been reformatted somewhat in terms of heading numbering, page numbers, and so on, to bring them into line with the remainder of this document. Secondly, the actual guidelines also contain a "ready reference" listing of the tags. Since this is identical to the listing given (with interspersed commentary) in the first part of the document, I have not repeated it.

The content of the guidelines largely recapitulates information given in Chapters 3 and 4 of this thesis, but there are also some details that are not mentioned there, and of course many more examples of the various categories. The text is written with the intended audience of an informant performing manual tagging, and the style has kept as simple and lucid as possible. The guidelines are in keeping with the U1 version of the tagset, since that was the one used by the manual taggers.

The tagging guidelines are divided into two broad sections: a description of the tagset (starting on the next page), and a section of guidelines discussing "how to tag" and some problematic areas (starting on page 449).

Tagset Definition: the U1 Tagset

The following is a description of the U1 tagset for use with Urdu texts and corpora. In the tables that list the tags, the first column gives the description of the tag, the second column gives an example of a word that would take that tag if one is deemed necessary, and the third column gives the tag. For tags that can only apply to a very few words, all the words it can apply to are listed. There are less than 400 tags.

Most Urdu words in this document are written in Perso-Arabic characters. However, some have been written in Latin characters. The transliteration system used should be fairly straightforward, but a full description of it can be provided if necessary.

A4.1 Nouns

Tags for nouns begin with the letter **N**, for example **N**NMM1N, **N**PUF1O.

A4.1.1 Common versus proper nouns

A proper noun is a "name" – i.e. it refers to one or more particular things, people or places, for instance کراچی. A common noun is not a name – it refers to one or more members of a class of things, for instance لڑکا. The second character of the tag indicates whether a noun is common (**N**) or **P**roper (e.g. N**P**UM1N, N**N**MM1N).

You can have two proper nouns in a row quite commonly – e.g. a person's first name and surname, or place names like "New Zealand" or *islām ābād*.

A4.1.2 Gender

Nouns can be marked or unmarked, and masculine or feminine. Whether a noun is "marked" or "unmarked" has to do with what kind of affixes get added to it to form the plural. For example:

روپیہ	rūpayah	(marked masculine)
گھر	ghar	(unmarked masculine)
بچّی	baccī	(marked feminine)
کتاب	kitāb	(unmarked feminine)

Marked masculine nouns change the –ā ending to –ē to form the plural. Marked feminine nouns change the –ī / –iyā ending to –iyā~. Unmarked masculine nouns stay the same in the nominative plural, and unmarked feminine nouns add the ending –ē~. Nouns which follow a pluralisation pattern borrowed from another language (e.g. Arabic "broken" plurals, or English plurals ending in –s) are also treated as unmarked.

The ending on the singular is not a 100% reliable guide because there are some nouns ending in –ā and –ī which are not marked nouns (often these words are borrowed from Persian, Arabic or Sanskrit – lots of names are like this). However the relationship between a noun and its plural will always tell you whether it is marked or unmarked.

A small number of nouns have both marked and unmarked plurals – for example, the originally Arabic noun *masaalah* has a marked (NNMM2N) plural *maslē* and an unmarked (NNUM2N) plural *masāil*. The singular should be treated as

marked (i.e. *masaalah* NNMM1N).

The **M**arked / **U**nmarked distinction is indicated in the third character of the tag. So: NN**M**M1N is marked, NN**U**M1N is not. The fourth character of the tag indicates the gender: NNM**M**1N is **M**asculine, NNM**F**1N is **F**eminine.

Some proper nouns (names) can be applied to both men and women (e.g. surnames). Nouns like this should be tagged as masculine if they refer to a man, and feminine if they refer to a woman.

A4.1.3 Number

Nouns can be singular (e.g. روپیہ) or plural (e.g. روپئے). This is indicated by the fifth character of the tag, which is a 1 or a 2: NNMM**1**N is singular, NNMM**2**N is plural. Note that unmarked masculine nouns look the same in the plural as the singular, so you must decide from the context what the number is.

A4.1.4 Case

Nouns have three cases, **N**ominative, **O**blique and **V**ocative. Nominative is the "normal" case; oblique occurs before postpositions; vocative is used when the noun is being "addressed". Note that for many nouns some of the forms look the same, e.g. nominative plural and oblique singular.

لڑکا , لڑکے	laRkā, laRkē	(nominative singular/plural)
لڑکے , لڑکوں	laRkē, laRkō~	(oblique singular/plural)
لڑکے , لڑکو	laRkē, laRkō	(vocative singular/plural)

Case is indicated by N, O or V as the last letter of the tag: e.g. NNMM1**N**, NNMM1**O**, NNMM1**V**.

A4.1.5 The tags for nouns

Description	Example	Tag
Common marked masculine singular nominative noun	لڑکا	NNMM1N
Common marked masculine singular oblique noun	لڑکے	NNMM1O
Common marked masculine singular vocative noun	لڑکے	NNMM1V
Common marked masculine plural nominative noun	لڑکے	NNMM2N
Common marked masculine plural oblique noun	لڑکوں	NNMM2O
Common marked masculine plural vocative noun	لڑکو	NNMM2V
Common marked feminine singular nominative noun	چڑیا	NNMF1N
Common marked feminine singular oblique noun	چڑیا	NNMF1O
Common marked feminine singular vocative noun	چڑیا	NNMF1V
Common marked feminine plural nominative noun	چڑیاں	NNMF2N
Common marked feminine plural oblique noun	چڑیوں	NNMF2O
Common marked feminine plural vocative noun	چڑیو	NNMF2V
Common unmarked masculine singular nominative noun	بھائی	NNUM1N
Common unmarked masculine singular oblique noun	بھائی	NNUM1O
Common unmarked masculine singular vocative noun	بھائی	NNUM1V
Common unmarked masculine plural nominative noun	بھائی	NNUM2N
Common unmarked masculine plural oblique noun	بھائیوں	NNUM2O
Common unmarked masculine plural vocative noun	بھائیو	NNUM2V
Common unmarked feminine singular nominative	بہن	NNUF1N

noun		
Common unmarked feminine singular oblique noun	بہن	NNUF1O
Common unmarked feminine singular vocative noun	بہن	NNUF1V
Common unmarked feminine plural nominative noun	بہنیں	NNUF2N
Common unmarked feminine plural oblique noun	بہنوں	NNUF2O
Common unmarked feminine plural vocative noun	بہنو	NNUF2V
Proper marked masculine singular nominative noun[1]		NPMM1N
Proper marked masculine singular oblique noun		NPMM1O
Proper marked masculine singular vocative noun		NPMM1V
Proper marked masculine plural nominative noun		NPMM2N
Proper marked masculine plural oblique noun		NPMM2O
Proper marked masculine plural vocative noun		NPMM2V
Proper marked feminine singular nominative noun		NPMF1N
Proper marked feminine singular oblique noun		NPMF1O
Proper marked feminine singular vocative noun		NPMF1V
Proper marked feminine plural nominative noun		NPMF2N
Proper marked feminine plural oblique noun		NPMF2O
Proper marked feminine plural vocative noun		NPMF2V
Proper unmarked masculine singular nominative noun		NPUM1N
Proper unmarked masculine singular oblique noun		NPUM1O
Proper unmarked masculine singular vocative noun		NPUM1V
Proper unmarked masculine plural nominative noun		NPUM2N
Proper unmarked masculine plural oblique noun		NPUM2O
Proper unmarked masculine plural vocative noun		NPUM2V
Proper unmarked feminine singular nominative noun		NPUF1N

[1] Intentionally, no examples are given for proper nouns, since they exist in a great variety. Proper nouns which are marked seem to be very rare, but do exist (e.g. *Sabuhi*).

Proper unmarked feminine singular oblique noun		NPUF1O
Proper unmarked feminine singular vocative noun		NPUF1V
Proper unmarked feminine plural nominative noun		NPUF2N
Proper unmarked feminine plural oblique noun		NPUF2O
Proper unmarked feminine plural vocative noun		NPUF2V

A4.2 Verbs

There are 113 tags for verbs. All verbal tags begin with V. The second character of a verb tag indicates whether it is a lexical verb (V), a general auxiliary verb (X), or a special auxiliary verb (some other character).

A small group of lexical verbs can act as general auxiliary verbs. They have tags beginning in VX. They are discussed below.

The special auxiliary verbs have tags beginning with VR, VG, VC, VH, and so on. They are classes of one word or a few words only. They are discussed in a separate section at the end of the section on verbs.

A4.2.1 Lexical verbs

Lexical verbs occur in a number of inflected forms. The vast majority of verbs are lexical verbs. They have tags beginning in VV. The third character indicates which form the verb is in: 0 (zero) for root, N for infinitive, T for imperfective participle, Y for perfective participle, S for subjunctive, and I for imperative.

A4.2.1.1 The root

The root of the verb can occur without any affixes. Since this only has one form, there is only one tag.

Description	Example	Tag (Roman)
Root form lexical verb	سن	VV0

A4.2.1.2 The infinitive

Tags for verb infinitives start in VVN. Infinitives can be used as verbs or as verbal nouns. The last three characters show gender (4th character: **M** or **F**), number (5th character: **1** or **2**) and case (6th character: **N** or **O**). But you should note that "case" only applies when the infinitive is masculine singular. When the infinitive is feminine or masculine plural, the last character should be left out.

Description	Example	Tag
Infinitive lexical verb, masculine singular nominative	سننا	VVNM1N
Infinitive lexical verb, masculine singular oblique	سننے	VVNM1O
Infinitive lexical verb, masculine plural nominative	سننے	VVNM2
Infinitive lexical verb, feminine singular nominative	سننی	VVNF1
Infinitive lexical verb, feminine plural nominative	سننی	VVNF2

A4.2.1.3 The participles

Urdu has two participles, the imperfective and the perfective. These can occur with an auxiliary such as ہے after them, or alone. The perfective tags start in VVY and the imperfective tags in VVT.

Like adjectives, the tags for participles have characters for gender (4th character: **M** or **F**), number (5th character: **1** or **2**) and case (6th character: **N** or **O**).

There is an important difference between adjectives and participles. Adjectives have the same form for the feminine singular and feminine plural. While some feminine plural participles can end in ی like a feminine singular participle, some end in یں .

Description	Example	Tag
Masculine singular (nominative) imperfective participle lexical verb	سنتا	VVTM1N
Masculine singular oblique imperfective participle lexical verb	سنتے	VVTM1O
Masculine plural (nominative) imperfective participle lexical verb	سنتے	VVTM2N
Masculine plural oblique imperfective participle lexical verb	سنتے	VVTM2O
Feminine singular (nominative) imperfective participle lexical verb	سنتی	VVTF1N
Feminine singular oblique imperfective participle lexical verb	سنتی	VVTF1O
Feminine plural (nominative) imperfective participle lexical verb	سنتی , سنتیں	VVTF2N
Feminine plural oblique imperfective participle lexical verb	سنتی	VVTF2O
Masculine singular (nominative) perfective participle lexical verb	سنا	VVYM1N
Masculine singular oblique perfective participle lexical verb	سنے	VVYM1O
Masculine plural (nominative) perfective participle lexical verb	سنے	VVYM2N

Masculine plural oblique perfective participle lexical verb	سنے	VVYM2O
Feminine singular (nominative) perfective participle lexical verb	سنی	VVYF1N
Feminine singular oblique perfective participle lexical verb	سنی	VVYF1O
Feminine plural (nominative) perfective participle lexical verb	سنی , سنیں	VVYF2N
Feminine plural oblique perfective participle lexical verb	سنی	VVYF2O

A4.2.1.4 The subjunctive

The tags for subjunctive verbs start with VVS. They do not have a character for gender. Instead they have a character for "person" This works as follows: **I** am the first person, **you** are the second person, and **he/she/it** is the third person. Verbs that go with میں or ہم are tagged as first person, verbs with go with تو or تم are tagged as second person, and so on.

Verb forms which agree with آپ are tagged as third person plural forms in the subjunctive. آپ always agrees with the third person plural form of the verb.

The fourth letter of the subjunctive tags is based on the pronoun that the verb agrees with: **M** for میں, **T** for تو, or **V** for وہ . Note that the **M** here does *not* mean masculine! The fifth letter shows number (**1** for singular or **2** for plural, as before).

Description	Example	Tag
First person singular subjunctive lexical verb	سنوں	VVSM1
First person plural subjunctive lexical verb	سنیں	VVSM2
Second person singular subjunctive lexical verb	سنے	VVST1
Second person plural subjunctive lexical verb	سنو	VVST2

Third person singular subjunctive lexical verb	سنے	VVSV1
Third person plural subjunctive lexical verb	سنیں	VVSV2

A4.2.1.5 The imperative

The tags for the imperative (command or request forms) start in VVI. They end in **T1** for the form that goes with *tū*, **T2** for the form that goes with *tum*, and **A** for the form that goes with *āp*. Note that the VVIT1 form looks just like the root form (VV0), and the VVIT2 looks just like the second person plural subjunctive form (VVST2).

Description	Examples	Tag
Second person singular imperative lexical verb	سن	VVIT1
Second person singular imperative lexical verb	سنو	VVIT2
Second person honorific imperative lexical verb	سنئے	VVIA

A4.2.2 General auxiliary verbs

A small number of verbs can be used as "general auxiliaries". General auxiliaries occur in verb phrases which have two verbs together, where neither verb is a "special auxiliary" (see below). The first is in the root form and carries most of the meaning. The second can be in any form and adds only a "shade" to the meaning of the first. This second verb is called a general auxiliary. Examples are جانا , دینا , سکنا .

All verbs that can be general auxiliaries can be lexical verbs as well, if they occur outside this two-verb phrase. These verbs are tagged with VV tags if they are lexical and VX tags if they are acting as general auxiliaries.

General auxiliaries are defined as those which follow a lexical verb *in its root form*. If you have a two-verb construction where the first verb (the one that carries most of the meaning) is not in the root form, then the second verb is *not* a general auxiliary.

There are two special cases:

- the verb جانا forms the passive when it is preceded by the perfective participle of a lexical verb. When جانا is used in this way, tag it as a general auxiliary.
- When the root of a lexical verb is followed by کر , tag کر as a general auxiliary. Note that only the root form کر can ever be a general auxiliary, and کرنا is not an auxiliary verb in any case other than this.

Most of the verbs which can act as general auxiliaries are listed below (other than کرنا). The list is still open. If you think you see an example of a verb not on this list acting as a general auxiliary, you should report the example and the verb may be added to the list.

saknā	سکنا	pānā	پانا
cuknā	چکنا	jānā	جانا
paRnā	پڑنا	nikalnā	نکلنا
uThnā	اُٹھنا	baiThnā	بیٹھنا
dēnā	دینا	lēnā	لینا
Dālnā	ڈالنا	rakhnā	رکھنا
guzarnā	گزرنا	calnā	چلنا

It is rare to have two general auxiliaries in a row but it can happen.

The tags for general auxiliary verbs are exactly the same as for lexical verbs, except that they start with VX instead of VV. A full list follows below.

Description	Example	Tag
Root form general auxiliary verb	پڑ	VX0
Infinitive general auxiliary verb, masculine singular nominative	پڑنا	VXNM1N
Infinitive general auxiliary verb, masculine singular oblique	پڑنے	VXNM1O
Infinitive general auxiliary verb, masculine plural nominative	پڑنے	VXNM2
Infinitive general auxiliary verb, feminine singular nominative	پڑنی	VXNF1
Infinitive general auxiliary verb, feminine plural nominative	پڑنی	VXNF2
Masculine singular (nominative) imperfective participle general auxiliary verb	پڑتا	VXTM1N
Masculine singular oblique imperfective participle general auxiliary verb	پڑتے	VXTM1O
Masculine plural (nominative) imperfective participle general auxiliary verb	پڑتے	VXTM2N
Masculine plural oblique imperfective participle general auxiliary verb	پڑتے	VXTM2O
Feminine singular (nominative) imperfective participle general auxiliary verb	پڑتی	VXTF1N
Feminine singular oblique imperfective participle general auxiliary verb	پڑتی	VXTF1O
Feminine plural (nominative) imperfective participle general auxiliary verb	پڑتیں , پڑتی	VXTF2N
Feminine plural oblique imperfective participle general auxiliary verb	پڑتی	VXTF2O
Masculine singular (nominative) perfective participle general auxiliary verb	پڑا	VXYM1N
Masculine singular oblique perfective participle general auxiliary verb	پڑے	VXYM1O
Masculine plural (nominative) perfective participle general auxiliary verb	پڑے	VXYM2N
Masculine plural oblique perfective participle general auxiliary verb	پڑے	VXYM2O
Feminine singular (nominative) perfective participle general auxiliary verb	پڑی	VXYF1N

Feminine singular oblique perfective participle general auxiliary verb	پڑی	VXYF1O
Feminine plural (nominative) perfective participle general auxiliary verb	پڑیں , پڑی	VXYF2N
Feminine plural oblique perfective participle general auxiliary verb	پڑی	VXYF2O
First person singular subjunctive general auxiliary verb	پڑوں	VXSM1
First person plural subjunctive general auxiliary verb	پڑیں	VXSM2
Second person singular subjunctive general auxiliary verb	پڑے	VXST1
Second person plural subjunctive general auxiliary verb	پڑو	VXST2
Third person singular subjunctive general auxiliary verb	پڑے	VXSV1
Third person plural subjunctive general auxiliary verb	پڑیں	VXSV2
Second person singular imperative general auxiliary verb	پڑ	VXIT1
Second person singular imperative general auxiliary verb	پڑو	VXIT2
Second person honorific imperative general auxiliary verb	پڑئے	VXIA

A4.2.3 Special auxiliary verbs

A4.2.3.1 gā

The form *gā* indicates future tense. The tag for *gā* begins with VG. The other characters indicate gender (4th character: **M** or **F**), number (5th character: **1** or **2**).

Description	Example	Tag (Roman)
Masculine singular future auxiliary *gā*	گا	VGM1
Masculine plural future auxiliary *gē*	گے	VGM2
Feminine singular future auxiliary *gī*	گی	VGF1

Feminine plural future auxiliary *gī*	گی	VGF2

A4.2.3.2 rahā

This auxiliary element indicates continuous or durative aspect of a verb. The tag for *rahā* begins with VR. The other characters indicate gender (4th character: **M** or **F**), number (5th character: **1** or **2**).

Do not confuse the auxiliary *rahā* with the (historically related) lexical verb *rahnā* – the perfective participle *rahā* looks the same but must be tagged differently. *rahā* is only an auxiliary when it occurs in a verb phrase like چائے پی رہی ہے . When it is not an auxiliary, *rahā* receives the tag VVYM1N, *rahī* receives VVYF1N or VVYF2N, and so on.

Description	Example	Tag
Masculine singular durative auxiliary *rahā*	رہا	VRM1
Masculine plural durative auxiliary *rahē*	رہے	VRM2
Feminine singular durative auxiliary *rahī*	رہی	VRF1
Feminine plural durative auxiliary *rahī*	رہی	VRF2

A4.2.3.3 cāhiē

The auxiliary *cāhiē* comes after the infinitive of a lexical verb. It is tagged differently to the historically related verb *cāhnā*.

Description	Example	Tag
Singular cāhiē-type auxiliary	چاہئے	VC1
Plural cāhiē-type auxiliary	چاہئیں	VC2

A4.2.3.4 hōnā

The verb *hōnā* is a very common auxiliary. It has special tags, which are always used for this verb. Some of the parts of *hōnā* are equivalent to the parts of lexical verbs. Their tags are the same for those of lexical verbs, except that they begin in VH instead of VV.

Description	Example	Tag
Root form *hō*	ہو	VH0
Infinitive *hōnā*, masculine singular nominative	ہونا	VHNM1N
Infinitive *hōnē*, masculine singular oblique	ہونے	VHNM1O
Infinitive *hōnē*, masculine plural nominative	ہونے	VHNM2
Infinitive *hōnī*, feminine singular nominative	ہونی	VHNF1
Infinitive *hōnī*, feminine plural nominative	ہونی	VHNF2
Masculine singular (nominative) imperfective participle *hōtā*	ہوتا	VHTM1N
Masculine singular oblique imperfective participle *hōtē*	ہوتے	VHTM1O
Masculine plural (nominative) imperfective participle *hōtē*	ہوتے	VHTM2N
Masculine plural oblique imperfective participle *hōtē*	ہوتے	VHTM2O
Feminine singular (nominative) imperfective participle *hōtī*	ہوتی	VHTF1N
Feminine singular oblique imperfective participle *hōtī*	ہوتی	VHTF1O
Feminine plural (nominative) imperfective participle *hōtī / hōtī~*	ہوتیں , ہوتی	VHTF2N

Feminine plural oblique imperfective participle *hōtī*	ہونی	VHTF2O
Masculine singular (nominative) perfective participle *hūā*	ہوا	VHYM1N
Masculine singular oblique perfective participle *hūē*	ہونے	VHYM1O
Masculine plural (nominative) perfective participle *hūē*	ہونے	VHYM2N
Masculine plural oblique perfective participle *hūē*	ہونے	VHYM2O
Feminine singular (nominative) perfective participle *hūī*	ہونی	VHYF1N
Feminine singular oblique perfective participle *hūī*	ہونی	VHYF1O
Feminine plural (nominative) perfective participle *hūī / hūī~*	ہونیں , ہونی	VHYF2N
Feminine plural oblique perfective participle *hūī*	ہونی	VHYF2O
First person singular subjunctive *hū~*	ہوں	VHSM1
First person plural subjunctive *hō~*	ہوں	VHSM2
Second person singular subjunctive *hō*	ہو	VHST1
Second person plural subjunctive *hō*	ہو	VHST2
Third person singular subjunctive *hō*	ہو	VHSV1
Third person plural subjunctive *hō~*	ہوں	VHSV2
Second person singular imperative *hō*	ہو	VHIT1
Second person plural imperative *hō*	ہو	VHIT2
Second person honorific imperative		VHIA

The irregular present tense and irregular past tense both require separate tags. The present tense is indicated by VHH. The fourth and fifth characters of the tags show person and number, as with the subjunctive.

Description	**Example**	**Tag**
First person singular indicative present *hū~*	ہوں	VHHM1
First person plural indicative present *hai~*	ہیں	VHHM2

Second person singular indicative present *hai*	ہے	VHHT1
Second person plural indicative present *hō*	ہو	VHHT2
Third person singular indicative present *hai*	ہے	VHHV1
Third person plural indicative present *hai~*	ہیں	VHHV2

The past tense is indicated by VHP. The other characters indicate gender (4th character: **M** or **F**), number (5th character: **1** or **2**).

Description	Examples	Tag
Masculine singular indicative past *thā*	تھا	VHPM1
Masculine plural indicative past *thē*	تھے	VHPM2
Feminine singular indicative past *thī*	تھی	VHPF1
Feminine plural indicative past *thī~*	تھیں	VHPF2

A4.3 Adjectives and similar words

Tags for adjectives and similar words such as numerals begin with J. The second character of the tag indicates the type of word: JJ or JP means an adjective, JD a determiner-like adjective such as a number or a fraction, and JX means non-lexical adjectival element.

A4.3.1 Adjectives

Adjectives have tags beginning in JJ. Marked adjectives – and some other words discussed below which work in the same way – have three characters after the JJ: indicating gender (3rd character: **M** or **F**), number (4th character: **1** or **2**) and case

(5th character: **N** or **O**). Vocative adjectives are the same as oblique adjectives and should be tagged as vocative case adjective (ending in –**O**).

The forms for marked feminine adjectives are always the same – i.e. they all end in –ī. However, depending on what sort of noun they are agreeing with, they can still be considered to be nominative or oblique, and singular or plural, and should be tagged accordingly.

Note that marked adjectives have no particular form for the vocative – the oblique form occurs with vocative nouns. So tag an adjective agreeing with a vocative noun as being oblique. The same applies to the other categories that follow the adjective pattern.

Unmarked adjectives (which are the same regardless of what number or gender they have) are marked as JJU. This category includes adjectives such as پسند which cannot modify a noun but only occur in the predicate of a sentence – i.e. cannot be used in front of a noun to modify that noun.

There are a total of 9 tags for "normal" adjectives.

Description	Example	Tag
Marked masculine singular nominative adjective	بڑا	JJM1N
Marked masculine singular oblique adjective	بڑے	JJM1O
Marked masculine plural nominative adjective	بڑے	JJM2N
Marked masculine plural oblique adjective	بڑے	JJM2O
Marked feminine singular nominative adjective	بڑی	JJF1N
Marked feminine singular oblique adjective	بڑی	JJF1O
Marked feminine plural nominative adjective	بڑی	JJF2N
Marked feminine plural oblique adjective	بڑی	JJF2O

Unmarked adjective	خطرناک	JJU

A4.3.2 Determiner-like adjectives

Determiner-like adjectives have tags that begin in JD-.

A4.3.2.1 Indefinite determiners

One set of determiner-like adjectives, the indefinite determiners, are just tagged with JD. The words in this class are as follows:

zyādah	more
kāfī	enough
har	every
kaī	several
tamām	all
kul	all
zarā	a little
cand	few
bahut	
nihāyat	(not certain about this one?)

Note that some of these words can also function as adverbs. They act as JD when they determine a noun, and as adverbs when they occur adverbially outside a noun phrase.

Some (like *bahut* and *zyāda*) can be JD or RD. They are JD when they

determine a following noun, and RD when they indicate the degree of a following adjective. See also below.

The JD class is a closed class, meaning that we do not expect there to be very many other words in this category: any that you find should be reported and will be added to this list.

Description	Example	Tag
Indefinite determiner	زیادہ ، کافی	JD

A4.3.2.2 Numerals

Numerals are numbers, either written out as words or written using figures (Arabic-Indic or Western). There are a total of 19 tags in this category.

There are three tags for cardinal numbers. The normal sort are tagged JDNU (since they are unmarked adjectives, there is no need to tag for gender or case). Cardinal numbers can also be oblique – with the same suffix as a plural unmarked noun – the tag for this is JDNUO. Some numerals have a shorter form (a "clitic") which occurs before the multiplicative suffix *gunā*. The tag for this is JDNUC. For example, چوگنا.(The "gunā" part should be tagged as if it were a separate word – see below.)

Ordinal numbers are marked adjectives, their final three characters indicate gender, number and case in the same way as marked adjectives do.

Description	Example	Tag
Cardinal number	ایک ، اٹھارہ	JDNU

Oblique cardinal number	دونوں	JDNUO
Pre-multiplicative clitic cardinal number *du–*, *ti–*, *cau–*	چو(گنا)	JDNUC
Masculine singular nominative ordinal number	تسرا	JDNM1N
Masculine singular oblique ordinal number	تسرے	JDNM1O
Masculine plural nominative ordinal number	تسرے	JDNM2N
Masculine plural oblique ordinal number	تسرے	JDNM2O
Feminine singular nominative ordinal number	تسری	JDNF1N
Feminine singular oblique ordinal number	تسری	JDNF1O
Feminine plural nominative ordinal number	تسری	JDNF2N
Feminine plural oblique ordinal number	تسری	JDNF2O

A4.3.2.3 Fractions

Words for fractions are given tags that begin in JDF. The following is the full list of fraction words:

پون ، پونا	سوا	آدھا ، آدھ
ساڑھے	ڈیڑھ	چوتھائی
تہائی	ڈھائی	

Some fraction words (*cauthāī, tihāī*) can also function as nouns, in which case they should receive standard noun tagging. Most fraction words are unmarked and receive the tag JDFU, but two (*paunā, ādhā*) are marked and have a final three characters that indicate gender, number and case in the same way as the tags for marked adjectives.

Description	Example	Tag
Unmarked fraction	سوا	JDFU
Masculine singular nominative fraction	پونا	JDFM1N
Masculine singular oblique fraction	پونے	JDFM1O
Masculine plural nominative fraction	پونے	JDFM2N
Masculine plural oblique fraction	پونے	JDFM2O
Feminine singular nominative fraction	پونی	JDFF1N
Feminine singular oblique fraction	پونی	JDFF1O
Feminine plural nominative fraction	پونی	JDFF2N
Feminine plural oblique fraction	پونی	JDFF2O

A4.3.3 "Y-V-K-J" determiners

The determiners that fall into the "Y-V-K-J" sets of words (see the section on Pronouns below) have tags beginning in JDY, JDV, JDK, and JDJ; their last three characters indicate gender, number and case in the same way as marked adjectives.

Note that the word *taisā* (*taisē*, etc.) which appears in some set phrases is tagged as JDV*** , just like *vaisā*.

Description	Example	Tag
Masculine singular nominative proximal demonstrative adjective (*itnā, aisā*)	اِتنا ، ایسا	JDYM1N
Masculine singular oblique proximal demonstrative adjective (*itnē, aisē*)	اِتنے ، ایسے	JDYM1O
Masculine plural nominative proximal demonstrative adjective (*itnē, aisē*)	اِتنے ، ایسے	JDYM2N
Masculine plural oblique proximal demonstrative adjective (*itnē, aisē*)	اِتنے ، ایسے	JDYM2O
Feminine singular nominative proximal demonstrative adjective (*itnī, aisī*)	اِتنی ، ایسی	JDYF1N

Feminine singular oblique proximal demonstrative adjective (*itnī, aisī*)	اِتنی ، ایسی	JDYF1O
Feminine plural nominative proximal demonstrative adjective (*itnī, aisī*)	اِتنی ، ایسی	JDYF2N
Feminine plural oblique proximal demonstrative adjective (*itnī, aisī*)	اِتنی ، ایسی	JDYF2O
Masculine singular nominative distal demonstrative adjective (*utnā, vaisā*)	اُتنا ، ویسا	JDVM1N
Masculine singular oblique distal demonstrative adjective (*utnē, vaisē*)	اُتنے ، ویسے	JDVM1O
Masculine plural nominative distal demonstrative adjective (*utnē, vaisē*)	اُتنے ، ویسے	JDVM2N
Masculine plural oblique distal demonstrative adjective (*utnē, vaisē*)	اُتنے ، ویسے	JDVM2O
Feminine singular nominative distal demonstrative adjective (*utnī, vaisī*)	اُتنی ، ویسی	JDVF1N
Feminine singular oblique distal demonstrative adjective (*utnī, vaisī*)	اُتنی ، ویسی	JDVF1O
Feminine plural nominative distal demonstrative adjective (*utnī, vaisī*)	اُتنی ، ویسی	JDVF2N
Feminine plural oblique distal demonstrative adjective (*utnī, vaisī*)	اُتنی ، ویسی	JDVF2O
Masculine singular nominative interrogative adjective (*kitnā, kaisā*)	کتنا ، کیسا	JDKM1N
Masculine singular oblique interrogative adjective (*kitnē, kaisē*)	کتنے ، کیسے	JDKM1O
Masculine plural nominative interrogative adjective (*kitnē, kaisē*)	کتنے ، کیسے	JDKM2N
Masculine plural oblique interrogative adjective (*kitnē, kaisē*)	کتنے ، کیسے	JDKM2O
Feminine singular nominative interrogative adjective (*kitnī, kaisī*)	کتنی ، کیسی	JDKF1N
Feminine singular oblique interrogative adjective (*kitnī, kaisī*)	کتنی ، کیسی	JDKF1O
Feminine plural nominative interrogative adjective (*kitnī, kaisī*)	کتنی ، کیسی	JDKF2N
Feminine plural oblique interrogative adjective (*kitnī, kaisī*)	کتنی ، کیسی	JDKF2O
Masculine singular nominative relative adjective (*jitnā, jaisā*)	جتنا ، جیسا	JDJM1N
Masculine singular oblique relative adjective (*jitnē, jaisē*)	جتنے ، جیسے	JDJM1O
Masculine plural nominative relative adjective (*jitnē, jaisē*)	جتنے ، جیسے	JDJM2N
Masculine plural oblique relative adjective (*jitnē, jaisē*)	جتنے ، جیسے	JDJM2O
Feminine singular nominative relative adjective (*jitnī, jaisī*)	جتنی ، جیسی	JDJF1N

Feminine singular oblique relative adjective (*jitnī, jaisī*)	جتنی ، جیسی	JDJF1O
Feminine plural nominative relative adjective (*jitnī, jaisī*)	جتنی ، جیسی	JDJF2N
Feminine plural oblique relative adjective (*jitnī, jaisī*)	جتنی ، جیسی	JDJF2O

A4.3.3.1 Non-lexical adjectival elements

The non-lexical adjectival elements have tags that start in JX. They are a closed class consisting of the various forms of *gunā* (JXG), *sā* (JXS) and *vālā* (JXV). The last three characters of these sets of tags indicates gender, number and case in the same way as marked adjectives.

Description	Example	Tag
Masculine singular nominative multiplicative marker *gunā*	گنا	JXGM1N
Masculine singular oblique multiplicative marker *gunē*	گنے	JXGM1O
Masculine plural nominative multiplicative marker *gunē*	گنے	JXGM2N
Masculine plural oblique multiplicative marker *gunē*	گنے	JXGM2O
Feminine singular nominative multiplicative marker *gunī*	گنی	JXGF1N
Feminine singular oblique multiplicative marker *gunī*	گنی	JXGF1O
Feminine plural nominative multiplicative marker *gunī*	گنی	JXGF2N
Feminine plural oblique multiplicative marker *gunī*	گنی	JXGF2O
Masculine singular nominative adjectival particle *sā*	سا	JXSM1N
Masculine singular oblique adjectival particle *sē*	سے	JXSM1O
Masculine plural nominative adjectival particle *sē*	سے	JXSM2N
Masculine plural oblique adjectival particle *sē*	سے	JXSM2O
Feminine singular nominative adjectival particle *sī*	سی	JXSF1N

Feminine singular oblique adjectival particle *sī*	سی	JXSF1O
Feminine plural nominative adjectival particle *sī*	سی	JXSF2N
Feminine plural oblique adjectival particle *sī*	سی	JXSF2O
Masculine singular nominative adjectival / occupational particle *vālā*	والا	JXVM1N
Masculine singular oblique adjectival / occupational particle *vālē*	والے	JXVM1O
Masculine plural nominative adjectival / occupational particle *vālē*	والے	JXVM2N
Masculine plural oblique adjectival / occupational particle *vālē / vālō~*[2]	والے	JXVM2O
Feminine singular nominative adjectival / occupational particle *vālī*	والی	JXVF1N
Feminine singular oblique adjectival / occupational particle *vālī*	والی	JXVF1O
Feminine plural nominative adjectival / occupational particle *vālī*	والی	JXVF2N
Feminine plural oblique adjectival / occupational particle *vālī*	والی	JXVF2O

A4.4 Pronouns

Pronouns have tags beginning in P. There are five groups of these: personal pronouns, personal possessive adjectives, "Y-V-K-J" pronouns, reflexive pronouns, and other pronouns.

A4.4.1 Personal pronouns

This category contains only the pronouns *mai~* and *tū*, and related forms such as their plural and oblique forms. Tags in this category begin with PP. The third

[2] Note that the alternative form for the masculine oblique plural, *vālō~*, does not occur in all the same contexts as *vālē*, and sometimes occurs in other, completely different contexts, being used as a type of noun. For example in the context *hamārē vālō~ kā*, "belonging to our people", *vālō~* is not JXV.

character indicates person (**M** or **T**, as for verbs – see above). The fourth indicates number (**1** or **2**), and the fifth indicates case (**N** or **O**).

Personal pronouns are not marked for gender: as with verbs, that which is marked for person is not marked for gender. (The "M" in the tags below signifies "first person", not "masculine".) They *are* marked for number and case.

Sometimes pronouns are used to refer to a different number of people than their "basic" number. For example, *ham*, which is basically plural, is sometimes used to refer to one person. In this case, the pronoun should still be tagged as plural – that is, *ham* is always PPM2N or PPM2O, even if it refers to just one person. Similarly, *tum* is always PPT2N or PPT2O, even if it refers to just one person.

In the same way, *mai~* and *tū* are always nominative. Even when they occur before the postposition *nē*, they should be considered nominative, even though a pronoun before a postposition is usually oblique. *mai~* and *tū* can never be oblique. *ham* and *tum*, which can also be oblique, should be tagged as nominative before *nē*.

Description	**Example**	**Tag**
First person singular nominative personal pronoun (*mai~*)	میں	PPM1N
First person singular oblique personal pronoun (*mujh*)	مجھ	PPM1O
First person plural nominative personal pronoun (*ham*)	ہم	PPM2N
First person plural oblique personal pronoun (*ham*)	ہم	PPM2O
Second person singular nominative personal pronoun (*tū*)	تو	PPT1N
Second person singular oblique personal pronoun (*tujh*)	تجھ	PPT1O
Second person plural nominative personal pronoun (*tum*)	تم	PPT2N
Second person plural oblique personal pronoun (*tum*)	تم	PPT2O

A4.4.2 Personal possessive adjectives

These are called “adjectives” because they take adjective marking, however they are classed as a type of pronoun because they correspond to the personal pronouns above. Their tags begin in PG. The third and fourth characters show the person (**M** for first person and **T** for second person, as for personal pronouns) and number (**1** or **2**) of the person referred to, i.e. the possessor.

Like adjective tags, their final three characters indicate gender, number and case of the noun being modified, i.e. the thing that is possessed.

Description	Example	Tag
First person singular masculine singular nominative possessive adjective (*mērā*)	میرا	PGM1M1N
First person singular masculine singular oblique possessive adjective (*mērē*)	میرے	PGM1M1O
First person singular masculine plural nominative possessive adjective (*mērē*)	میرے	PGM1M2N
First person singular masculine plural oblique possessive adjective (*mērē*)	میرے	PGM1M2O
First person singular feminine singular nominative possessive adjective (*mērī*)	میری	PGM1F1N
First person singular feminine singular oblique possessive adjective (*mērī*)	میری	PGM1F1O
First person singular feminine plural nominative possessive adjective (*mērī*)	میری	PGM1F2N
First person singular feminine plural oblique possessive adjective (*mērī*)	میری	PGM1F2O
First person plural masculine singular nominative possessive adjective (*hamārā*)	ہمارا	PGM2M1N
First person singular masculine singular oblique possessive adjective (*hamārē*)	ہمارے	PGM2M1O
First person singular masculine plural nominative possessive adjective (*hamārē*)	ہمارے	PGM2M2N
First person singular masculine plural oblique possessive adjective (*hamārē*)	ہمارے	PGM2M2O
First person singular feminine singular nominative possessive adjective (*hamārī*)	ہماری	PGM2F1N
First person singular feminine singular oblique	ہماری	PGM2F1O

possessive adjective (*hamārī*)		
First person singular feminine plural nominative possessive adjective (*hamārī*)	ہماری	PGM2F2N
First person singular feminine plural oblique possessive adjective (*hamārī*)	ہماری	PGM2F2O
Second person singular masculine singular nominative possessive adjective (*tērā*)	تیرا	PGT1M1N
Second person singular masculine singular oblique possessive adjective (*tērē*)	تیرے	PGT1M1O
Second person singular masculine plural nominative possessive adjective (*tērē*)	تیرے	PGT1M2N
Second person singular masculine plural oblique possessive adjective (*tērē*)	تیرے	PGT1M2O
Second person singular feminine singular nominative possessive adjective (*tērī*)	تیری	PGT1F1N
Second person singular feminine singular oblique possessive adjective (*tērī*)	تیری	PGT1F1O
Second person singular feminine plural nominative possessive adjective (*tērī*)	تیری	PGT1F2N
Second person singular feminine plural oblique possessive adjective (*tērī*)	تیری	PGT1F2O
Second person plural masculine singular nominative possessive adjective (*tumhārā*)	تمہارا	PGT2M1N
Second person singular masculine singular oblique possessive adjective (*tumhārē*)	تمہارے	PGT2M1O
Second person singular masculine plural nominative possessive adjective (*tumhārē*)	تمہارے	PGT2M2N
Second person singular masculine plural oblique possessive adjective (*tumhārē*)	تمہارے	PGT2M2O
Second person singular feminine singular nominative possessive adjective (*tumhārī*)	تمہاری	PGT2F1N
Second person singular feminine singular oblique possessive adjective (*tumhārī*)	تمہاری	PGT2F1O
Second person singular feminine plural nominative possessive adjective (*tumhārī*)	تمہاری	PGT2F2N
Second person singular feminine plural oblique possessive adjective (*tumhārī*)	تمہاری	PGT2F2O

A4.4.3 Third person pronouns/demonstratives, interrogative and relative pronouns and determiners (also called Y-V-K-J pronouns)

This class of pronouns consists of all those pronouns that fall into the parallel "y-v-k-j word sets". These classes contain a variety of pronouns that are of similar

form, the first letter indicating what set they belong to, (see below for examples):

- y or a vowel indicates the set of proximal ("near") demonstratives
- v or t indicates the set of distal ("far") demonstratives
- k indicates the set of interrogatives
- j indicates the set of relative words

The tags for these words begin with PY, PV, PK or PJ. The third character indicates number (1 or 2) and the fourth indicates case (N or O). There is an "extra" oblique case in the plurals of these pronouns, which appears solely before the postposition *nē*. This extra case is given **E** in the fourth position.

Description	**Example**	**Tag**
Singular nominative proximal demonstrative pronoun (*yah*)	یہ	PY1N
Singular oblique proximal demonstrative pronoun (*is*)	اِس	PY1O
Plural nominative proximal demonstrative pronoun (*yah*)	یہ	PY2N
Plural oblique proximal demonstrative pronoun (*in*)	اِن	PY2O
Plural oblique proximal demonstrative pronoun before *nē* (*inhō~*)	اِنہوں	PY2E
Singular nominative distal demonstrative pronoun (*vah*)	وہ	PV1N
Singular oblique distal demonstrative pronoun (*us*)	اُس	PV1O
Plural nominative distal demonstrative pronoun (*vah*)	وہ	PV2N
Plural oblique distal demonstrative pronoun (*un*)	اُن	PV2O
Plural oblique distal demonstrative pronoun before *nē* (*unhō~*)	اُنہوں	PV2E
Singular nominative interrogative pronoun (*kyā, kaun*)	کیا، کون	PK1N

Singular oblique interrogative pronoun (*kis*)	کس	PK1O
Plural nominative interrogative pronoun (*kyā, kaun*)	کیا، کون	PK2N
Plural oblique interrogative pronoun (*kin*)	کن	PK2O
Plural oblique interrogative pronoun before *nē* (*kinhō~*)	کنہوں	PK2E
Singular nominative relative pronoun (*jō*)	جو	PJ1N
Singular oblique relative pronoun (*jis*)	جس	PJ1O
Plural nominative relative pronoun (*jō*)	جو	PJ2N
Plural oblique relative pronoun (*jin*)	جن	PJ2O
Plural oblique relative pronoun before *nē* (*jinhō~*)	جنہوں	PJ2E

A4.4.4 Reflexive pronouns

There are separate tags for reflexive pronouns, the reciprocal pronoun, and the reflexive possessive adjectives. Tags for reflexive possessive adjectives have characters indicating gender, number and case after the PGR tag: this works in the same way as for the personal possessive adjectives.

Description	Example	Tag
Reflexive pronoun (*āp, xud*)	آپ , خود	PRF
Reciprocal pronoun (*āpas*)	آپس	PRC
Masculine singular nominative reflexive possessive adjective (*apnā*)	اپنا	PGRM1N
Masculine singular oblique reflexive possessive adjective (*apnē*)	اپنے	PGRM1O
Masculine plural nominative reflexive possessive adjective (*apnē*)	اپنے	PGRM2N
Masculine plural oblique reflexive possessive adjective (*apnē*)	اپنے	PGRM2O
Feminine singular nominative reflexive possessive adjective (*apnī*)	اپنی	PGRF1N
Feminine singular oblique reflexive possessive	اپنی	PGRF1O

adjective (*apnī*)		
Feminine plural nominative reflexive possessive adjective (*apnī*)	اپنی	PGRF2N
Feminine plural oblique reflexive possessive adjective (*apnī*)	اپنی	PGRF2O

A4.4.5 Other pronouns

In this miscellaneous group of pronouns are included two indefinite pronouns, *kōī* and *kuch*, which are tagged as PNN; their oblique forms are tagged as PNO. Also included in the PNN category is *sab*, which has an inflected oblique plural which is tagged as PNO. One final pronoun is the honorific, respectful, or polite pronoun *āp*, which has its own tag PA. Note that this is different to the reflexive pronoun *āp*! The reflexive pronoun PRF refers back to something or someone mentioned earlier in the sentence, whereas the honorific pronoun PA is used as a more polite equivalent for *tum*.

Description	Example	Tag
Nominative indefinite pronoun (*kōī, kuch, sab*)	کوئی, کچھ	PNN
Oblique indefinite pronoun (*kīsī*, *kuch, sabhō*~)	کسی	PNO
Honorific pronoun (*āp*)	آپ	PA

A4.5 Articles

Some phrases borrowed from Arabic contain the Arabic definite article, which receives the single tag AL.

Description	Example	Tag
Arabic definite article	ال	AL

A4.6 Adverbs

There are lexical and non-lexical adverbs, which will be considered in turn. There are a total of 13 adverb tags. All begin with R.

A4.6.1 Lexical adverbs (RR)

There are two sorts of lexical adverbs: adverbs which are derived from adjectives by inflecting them to their masculine oblique form or adding a Persian or Arabic loaned derivational affix[3] (RRJ), and adverbs which are not (RR). Note that the RRJ form therefore often looks the same as a JJM1O form.

Many words can be both JJU and RR (for instance, *thīk*).

Description	Example	Tag
General adverb	ہمیشہ	RR
General adverb derived from adjective	بائیں	RRJ

[3] Words which may function as both adjectives and adverbs without any morphological modification (e.g. *pās*) would be marked with either an adjective tag or RR; the RRJ tag is not used for such words.

A4.6.2 Non-lexical adverbs

A4.6.2.1 Degree adverbs

Degree adverbs occur before adjectives, and modify the degree to which they apply. They are adjectives which indicate "very" or "more" or a similar notion when they occur before an adjective. These include *bahut*, *zarā*, *zyādah,* and *kāfī*[4]. Another example is *nihāyat*. This is a closed class, but if you find another word that you think should be on the list, please report it.

Description	Example	Tag
Degree adverb	زیادہ	RD

A4.6.2.2 Modal adverbs

There are a number of modal adverbs, including: *śāyad*, *zarūr*, *bhī*, *phir* and *sirf*. This is another closed class which you can add to if you find another one. There is also a subgroup of negative modal adverbs (word meaning "not").

Description	Example	Tag
Modal adverb	ضرور	RM
Negative modal adverb (*nahī*~, *nah*, *mat*)	نہیں، نہ، مت	RMN

[4] These are also indefinite determiners (see above).

A4.6.2.3 Y-V-K-J adverbs

There are also a number of adverbs in the y-v-k-j sets of words, with meanings of time, place and manner such as “now”, “then”, etc. This includes all interrogative and relative adverbs. Adverbs derived from adjectives of the y-v-k-j sets (see above) have separate tags.

Description	Example	Tag
Proximal demonstrative adverb (*ab*, *yahā~*, *idhar*, *yū~*)[5]	اب	RY
Proximal demonstrative adverb derived from adjective (*aisē)*	ایسے	RYJ
Distal demonstrative adverb (*tab*, *vahā~*, *udhar*, *tyū~*)	وہاں	RV
Distal demonstrative adverb derived from adjective (*vaisē)*	ویسے	RVJ
Interrogative adverb (*kab*, *kahā~*, *kidhar*, *kyō~*)	کیوں	RK
Interrogative adverb derived from adjective (*kaisē)*	کیسے	RKJ
Relative adverb (*jab*, *jahā~*, *jidhar*, *jū~*)	جدھر	RJ
Relative adverb derived from adjective (*jaisē)*	جیسے	RJJ

A4.7 Adpositions

There are 10 tags for adpositions. All begin in I.

There are some prepositions, mostly borrowed from Persian or Arabic. Otherwise there are many postpositions, such as تک. There is also a single postposition, *kā*, which varies for gender, number and case, and thus has an extra

[5] *yahā~, vahā~, kahā~* and *jahā~* shorten to *yah-*, *vah-*, etc., when followed by the clitic exclusive emphatic particle.

three characters on its tag which work in the same way as the corresponding characters in the tags for marked adjectives.

Description	Example	Tag
Preposition	از، فی	IB
Unmarked postposition	پر، میں	II
Marked masculine singular nominative postposition *kā*	کا	IIM1N
Marked masculine singular oblique postposition *kē*	کے	IIM1O
Marked masculine plural nominative postposition *kē*	کے	IIM2N
Marked masculine plural oblique postposition *kē*	کے	IIM2O
Marked feminine singular nominative postposition *kī*	کی	IIF1N
Marked feminine singular oblique postposition *kī*	کی	IIF1O
Marked feminine plural nominative postposition *kī*	کی	IIF2N
Marked feminine plural oblique postposition *kī*	کی	IIF2O

A4.8 Conjunctions

There are three types of conjunctions: simple coordinating, correlative coordinating, and subordinating. These are again closed list categories.

Correlative conjunctions (such as *yā…yā*) come in pairs. The first of the two should be tagged CCC and the second of the pair as CC. Note that many of the correlative conjunctions have another use as well, as a coordinating conjunction or as something else.

Note that phrases involving the relative (J-)set of pronouns, adjectives and adverbs are often translated into English by conjunctions, but are not to be tagged as conjunctions.

As a general rule, any conjunction ending in "کہ" should be considered a subordinating conjunction.

Description	Example	Tag
Coordinating conjunction	مگر	CC
Correlative coordinating conjunction	یا	CCC
Subordinating conjunction	کہ	CS

A4.9 Interjections

There is one tag for all interjections. Some examples: ہاں، ارے، واہ، ہائے، او.

Description	Example	Tag
Interjection	واہ	AU

A4.10 Unique

The Unique categories are single words that don't fit anywhere else. Each has its own tag.

The question marker *kyā* looks the same as the interrogative pronoun *kyā*, but the two should be tagged differently.

The izāfat – the "e" sound found linking together two nouns in phrases such as دریائے سندھ – is not always written. If it is not written in the phrase you are tagging, you should "pretend it isn't there", even if you think it ought to be. If it is

written in the text, however, you should separate it from the word it is attached to, and give it the ZZ tag.

Sentence tag words conclude a sentence and comment on it. They are not grammatically linked to the sentence. Common tag words include *sahī* and *nā*. They are tagged TT.

Each emphatic particle each has a tag of their own.

The Persian compound-forming conjunction is found in expressions such as روز و شب. The word before it and the word after it should each receive a separate tag.

There are also two tags for clitic versions of words that have other full tags listed above. Clitics should be separated from their host word during tagging and given their own tags. Clitics are given the same tag as their full-length forms, but with an appended C.

In the case of one set of YVKJ adverbs, the clitic version of the particle *hī* fuses with them too completely for them to be separated out and given separate tags; therefore these fused words are given the special unique tags given here.

The use of the LL tag is explained in the more detailed tagging guidelines document accompanying this one.

The IV tag is used for one very particular use of *kē* – when it is used following a verb root to create an alternative version of the structure of [verb root + *kar*] described above. (Some grammarians have called this a "conjunctive participle"). So for example, *sun kar* should be tagged VV0 VX0 as already described; however, the alternative structure *sun kē* (with the same meaning) should be tagged VV0 IV.

Description	Example	Tag
Question marker *kyā*	کیا	QQ
izāfat	نے	ZZ
Sentence tag-word	سہی	TT
Clitic postposition[6] *ē*, *ē*~, *hē*~	ے، یں، ہیں	IIC
Contrastive emphatic particle *tō*	تو	XT
Exclusive emphatic particle *hī*	ہی	XH
Clitic exclusive emphatic particle *ī*, *ī*~, *hī*~	ی، یں، ہیں	XHC
Persian compound-forming conjunction *ō*	و	OO
Nongrammatical lexical element	زمہ	LL
Verbal postposition *kē*	کے	IV
Fused proximal demonstrative adverb and exclusive emphatic particle: *yahā*~ + *hī* = *yahī*~	یہیں	RYXHC
Fused distal demonstrative adverb and exclusive emphatic particle: *vahā*~ + *hī* = *vahī*~	وہیں	RVXHC
Fused interrogative adverb and exclusive emphatic particle: *kahā*~ + *hī* = *yahī*~	کہیں	RKXHC
Fused relative adverb and exclusive emphatic particle: *jahā*~ + *hī* = *jahī*~[7]	جہیں	RJXHC

A4.11 Residual

The 8 "residual" tags are for anything in the text that isn't really a word: foreign words, things written in other alphabets, mathematical formulae, etc.

The FF tag for "foreign words" is meant to cover words or phrases from other languages written in the Urdu alphabet. It is *not* meant to cover the large number of Persian, Arabic and English loanwords that exist in Urdu as an actual part of the Urdu language. The FX tag for "non-Perso-Arabic string" is for foreign words in other

[6] A form of *kō* added to a pronoun.

[7] *jahī*~ has been described as obsolete, so this tag would only be used in old-fashioned texts.

alphabets, or for other non-Perso-Arabic incursions into the text. FZ is for when a single character, not joined to any other, is being used as the name for that letter of the alphabet. FS is for non-alphabetic symbols (such as ☺ , which is nowadays quite common in texts from the internet). FA and FB are for acronyms and abbreviations. FU is for anything that can't be put into any of these categories. You should try to avoid using it whenever possible.

Description	Example	Tag
Foreign word		FF
Non-Perso-Arabic string		FX
Formula (e.g. mathematical)		FO
Letter of the alphabet		FZ
Other symbol		FS
Acronym		FA
Abbreviation		FB
Other unclassifiable non-Urdu element		FU

A4.12 Punctuation

The tag for a punctuation mark is usually the same as the punctuation mark.. There are 13 tags in this section.

Description	Example	Tag
Full stop (U+06D4)	۔	.
Comma (U+060C)	،	,
Question mark (U+061F)	؟	?

Exclamation mark (U+0021)	!	!
Colon (U+003A)	:	:
Semi-colon (U+061B)	؛	;
Neutral quotation mark (U+0022)	"	"
Open parenthesis (U+0028)	)	(
Close parenthesis (U+0029)	(	)
Open square bracket (U+005B)	[	[
Close square bracket (U+005D)	[	]

Some of the tags look as if they are the opposite way around to the actual punctuation mark ("mirrored") – this is because the tags are in the Latin alphabet, which goes in the opposite direction to the Urdu alphabet. That is, the Urdu opening bracket ")" should be tagged with the Latin opening bracket "(".

The tag for the neutral quotation mark should be used for both single and double quotation marks.

Tagging guidelines for the U1 tagset

A4.13 How to tag a word

A4.13.1 Basics of tagging

A line from a tagging file looks something like this:

s00098 w017 ہے	AJH

The number after "s" indicates the number of the sentence in the file. The number after "w" indicates the number of the word in the sentence. You can mostly ignore these numbers. The Urdu word follows. Then comes a blank space (actually a TAB character). At the end of the line are three letters that are used to indicate who has done the tagging – in files that are sent to you to tag, this will have been adjusted to a shortened version of your name. (In the example above, it is AJH for "Andrew Hardie"). Then, after this, there is a space.

To tag a word, simply move your cursor to the very end of the line – i.e. after the space that follows the shortened name – and type the appropriate letters and numbers. In this case you type VHHV1 and get a line that looks like this:

s00098 w017 ہے	AJH VHHV1

A4.13.2 Hints on speeding up tagging

You may wish to use CAPS LOCK since most of the tags are just capital letters.

If you find that you are typing a particular string of letters over and over again, try copying that string to the clipboard and simply using CTRL+V to insert it as necessary.

A4.13.3 SGML and the NULL tag

The files you are to tag contain SGML elements – text information contained in <angle brackets>. SGML elements do not receive tags – so they have automatically been labelled NULL (this means "no tag needed"). You can just pass over them as if they were not there.

A4.13.4 Typing errors in the original file

If a word has been mistyped in the file you are tagging – i.e. a letter has been typed incorrectly and it has made the word unclear – you should make a note of this, along with its s-number and its w-number. There are two approaches you can take to tagging words that have typos. If it is fairly clear from the context what word it should be, then you can tag it as if it was that word. If it is not entirely clear what word it should be, then tag it as FU.

A4.14 Problems arising from Ambiguity

A4.14.1 Ambiguous suffixes

For nouns, adjectives and verbs, you will often find that a single suffix has more than one usage. For example, the suffix *–ē* is used for masculine plural nouns, masculine singular oblique nouns, masculine adjectives, and a number of verbal forms.

You must tag for function, not for form – that is, you must tag according to what the word ending is doing, not according to what it is. So the same word-form may in one context be a masculine plural nominative adjective, and in another context a masculine singular oblique adjective.

A4.14.2 Ambiguous words

However, it is not just affix confusion that makes more than one tag possible for some words. For example, e.g. میں is tagged PPM1N when it is pronounced *mai~* and II when it is pronounced *me~*.

A4.15 Problems relating to nouns

A4.15.1 Case in unmarked nouns

Unmarked singular nouns, and marked feminine singular nouns as well, are the same in both nominative and oblique cases. So how do you know whether to tag them as

nominative or oblique?

Basically, you should give them the tag of the form that they would have got if they were marked nouns. So in any given context, you should think: if there was a marked noun here instead (e.g. *laRkā* or *kamrah*) would it have been nominative or oblique? Then tag the unmarked noun accordingly. As a general rule, nouns that are followed by a postposition (or governed by a postposition, as some theories of grammar put it) are very often oblique, although there are obviously exceptions.

A4.16 Problems relating to verbs

A4.16.1 Verb agreement for gender and number

Sometimes, the verb agrees with the subject and sometimes it agrees with its object. You must tag it the same, whichever it does – i.e. if a verb is feminine (e.g. VVYF1N) it doesn't matter whether it is agreeing with the subject or the object. Sometimes, the verb agrees with neither the subject or the object (e.g. if both the subject and the object are followed by a postposition like *nē* or *kō*). In this case, the verb will default back to the masculine singular form and should be tagged as such; e.g., VVTM1N or VXYM1N.

The pronoun *āp* does not have any verb agreement of its own. It usually takes the same verb forms as the third person plural. Verbs agreeing with *āp* should be tagged as third person plural verb forms. Similarly, if a plural verb is used with a singular third-person subject (as an honorific usage) then this should also be tagged as a plural

verb. Give tags based on the morphological number of the verb, not based on the semantic/pragmatic usage of the verb.

A4.16.2 cāhiē

The agreement rules for *cāhiē* are a bit complex. Firstly, *cāhiē* agrees with its object. If it occurs in the form *cāhiē~*, then it is VC2 (plural). However, if it is in the form *cāhiē*, then it is not necessarily singular since *cāhiē* will occur with a plural object if *cāhiē* is followed by a plural-marked auxiliary. So if the object is plural, *cāhiē* is VC2; if the object is singular, *cāhiē* is VC1. There is an exception to this as well: if the object is followed by a postposition, e.g. *kō*, then *cāhiē* is VC1 no matter what number the object is.

A4.16.3 rahā *versus* rahnā

This guideline is an expansion on what is said about the VR tags (VRM1, VRM2, VRF1, VRF2) in the main tagset document.

The VR tags are for use when *rahē*, *rahā*, or *rahī* are used as the durative auxiliary: that is, preceded by the verb root and followed by some form of *hōnā*: e.g. *vah kar rahā hai*, "he is doing", as contrasted with *vah kartā hai*, "he does". Any form of *rahnā* used anywhere other than in this very specific context is not tagged using the VR tags. Instead, the normal system for verbs should be used – the tags beginning in VV.

For example, in the phrase *rahē gā* meaning “will stay” or “will keep on (doing something)”, *rahē* should be tagged as VVSV1. Another example is the phrase *xuś rahā karē~* meaning “stay happy” or “remain happy”, which should be tagged JJU VVYM1N VVSV2. In these examples, *rahā* is not functioning as the durative auxiliary, but as part of *rahnā*.

A4.16.4 Participles

Participles can function as adjectives. However, they should be tagged using the VVY*** and VVT*** tags even when they are functioning as adjectives.

A4.16.5 Infinitives

An infinitive ending in *–ē* that occurs before a postposition or before *vālā* is always considered to be VVNM1O, never VVNM2.

A4.17 Problems relating to pronouns

A4.17.1 Number in pronouns

You should tag pronouns for their grammatical number, not their pragmatic number. For example, *tum* is often used pragmatically to refer to a single person in a way that is less impolite than the singular *tū*. However, it is always grammatically plural and should be tagged as such (PPT2N or PPT2O).

A4.17.2 *mai~* and *tū* followed by *nē*

Usually, a word before a postposition is in the oblique case, but there is an exception when *mai~* and *tū* are followed by *nē*. In this case, the pronouns take the shape of the nominative case, and should therefore be tagged as nominative case pronouns (PPM1N and PPT1N).

A4.17.3 *in* and *un*, *is* and *us*

These pairs of words have different tags, but in the Perso-Arabic alphabet look exactly the same if the vowel marks are not written in. Try to tell from context whether a *PV* tag or a *PY* tag is appropriate if this occurs.

A4.17.4 Difference between kyā_QQ and kyā_PK1N

The word *kyā* has two analyses. As an interrogative pronoun *kyā* (PK1N) refers to an unknown entity that is the focus of a question, which is expected to be specified in the answer. As a question marker *kyā* (QQ) does not refer to anything at all, it just lets you know that the sentence is a question. So in the sentence *kyā yah sahī hai?* ("is it true?") *kyā* would be QQ; but in the sentence *yah kyā hai?* ("what is this?") *kyā* would be PK1N.

A4.18 Other problematic areas

A4.18.1 Compound postpositions and *kē liyē*

There are many compound postpositions consisting of *kē* followed by another word, making up a compound postposition. In this type of structure, tag *kē* as normal, and tag the word after it as its own normal part of speech – this will usually be tagged as an adverb or an oblique noun. As a rule of thumb, if the second word occurs elsewhere as a noun, then tag it as an oblique noun; if it occurs elsewhere as an adverb, then tag it as an adverb.

Note in particular the compound postposition *kē liē* or *kē liyē*. This should be tagged as IIM1O followed by RR (i.e. *liyē* is treated as an adverb even though it doesn't occur alone as an adverb). Note also that **whenever *kē* agrees with an adverb, it should be tagged IIM1O**.

This phrase, *kē liē*, is sometimes written as a single word – in which case, before tagging, it should be split, using the same method described for splitting off clitics (see the section on clitics below).

A4.18.2 The use of LL

The LL tag is used with single words – usually borrowed from Persian – which have a space in the middle, often between the root and an affix; *zimmah dār* is an example of such a word. Note that there is a whole set of affixes that can do this – but they don't

always do it (e.g. compare *samajhdār*). Since there is a space, each part has to have a tag. The tag that would be given to the whole word, if it was a single word, is applied to the second part of the word, and the first part receives the tag LL.

This LL, the "nongrammatical lexical element", is applied thus:

zimmah_LL dār_JJU

samajhdār_JJU

xūb_LL tarīn_JJU

kamtarīn_JJU

ham_LL zāt_JJU

hamdardī_NNUF1N

Note that LL is only to be used if *at least* one of the elements is incapable of standing as a word on its own. The typical example of this is the word *zimmah dār*, where the second element is not capable of standing on its own – it's really a suffix rather than an independent word – and so the JJU tag for the entire thing goes on *dār* and *zimmah* gets the LL tag.

If a single compound word is written with a space in the middle, but both parts are capable of standing alone as independent words, then the LL tag should not be used. Instead, give each word a separate tag (i.e. treat the whole thing as a phrase).

A4.18.3 Tagging words which are joined by *ō*

The Persian compound-forming conjunction is found in expressions such as روز و شب. It is tagged as OO, and the word before it and the word after it should each receive a separate tag – i.e. tag those words as normal.

Sometimes the words joined by *ō* do not occur on their own in Urdu, for example in the expression *tag ō dav*. In this case, give to each of the two words the tag that you would normally give to the compound as a whole.

A4.18.4 Reduplicated words

When a word is reduplicated (i.e. doubled up, or repeated as an "echo", sometimes with a change of vowel), the second word receives the same tag as the first. Examples include *thīk thāk*.

A4.18.5 Clitics

Clitics should always be tagged separately from the word they are attached to. If the tokenisation program has left them on a single line, you will need to split the line. to make both the new lines exactly the same as the lines produced by computer before adding the POS tags. For instance, if the computer gives you this…

s00022 w005 السلامُ	xxx

… then you should copy-and-paste the entire line, and then delete the "al" from the

second line and the “salāmu” from the first:

s00022 w005 ال	xxx
s00022 w005 سلامُ	xxx

You can then tag both words (the tags are in this case AL and NNUM1N). The same applies to any other word written as a single unit which has parts that require separate tags.

Another example would be the word *isī* (اِسی) which would need to be split into its component parts and tagged as اِس PY1O and ی XHC (clitic form of *hī*). Note also the similar pronoun-clitic combinations وہی ، یہی، اُسی ، ہمیں ، اِنہیں، مجھی and so on, which must be split up in the same way.

The same procedure should be used for Persian-derived compounds with “ō” in the middle – only in this case it must be divided into three lines (*word, ō, word*). You may also need to split a word if the typist has made an error and typed two words without a space between them.

A4.18.6 Greetings

Tagging greetings can be problematic. The Muslim greeting *alsalām ’alaikum* should be tagged as AL–NNUM1N–FF, with the *al* separated off from *salām*. The Hindu greetings *namaskār* and *namastē* should be tagged as AU when they are used as greetings, and as masculine unmarked nouns when they are used within the grammar of the sentence (e.g. as the subject or object of a verb). The English loan-word *hello* should be tagged as AU when used as a greeting, and as a masculine unmarked noun

when used within the grammar of the sentence.

A4.19 Individual problematic words and phrases

Word or phrase	*Tag as*
vaγairah	RR
bād mē~	RR II
kyō~kah	CS
xūs āmdēd	LL AU or LL NNUM1N (rarer)
dōbārah	RR
Tī vī	NN

(initials in English spelt out in Urdu, e.g. بی بی سی , should normally be tagged as FB).