Part-of-Speech Annotation Challenges in Marathi

Gajanan Rane, Nilesh Joshi, Geetanjali Rane, Hanumant Redkar,
Malhar Kulkarni and Pushpalk Bhattacharyya
Center For Indian Language Technology
Indian Institute of Technology Bombay, Mumbai, India
\{gkrane45, joshinilesh60, geetanjaleerane, hanumantredkar, 
malharku and pushpakkbh\}@gmail.com

Abstract
Part of Speech (POS) annotation is a significant challenge in natural language processing. The paper discusses issues and challenges faced in the process of POS annotation of the Marathi data from four domains viz., tourism, health, entertainment and agriculture. During POS annotation, a lot of issues were encountered. Some of the major ones are discussed in detail in this paper. Also, the two approaches viz., the lexical (L approach) and the functional (F approach) of POS tagging have been discussed and presented with examples. Further, some ambiguous cases in POS annotation are presented in the paper.

Keywords: Marathi, POS Annotation, POS Tagging, Lexical, Functional, Marathi POS Tagset, ILCI

1 Introduction
In any natural language, Part of Speech (POS) such as noun, pronoun, adjective, verb, adverb, demonstrative, etc., forms an integral building block of a sentence structure. POS tagging is one of the major activities in Natural Language Processing (NLP). In corpus linguistics, POS tagging is the process of marking/annotating a word in a text/corpus which corresponds to a particular POS. The annotation is done based on its definition and its context i.e., its relationship with adjacent and related words in a phrase, sentence, or paragraph. A simplified form of this is commonly taught to school-age children, in the identification of words as nouns, verbs, adjectives, adverbs, etc. The term ‘Part-of-Speech Tagging’ is also known as POS tagging, POST, POS annotation, grammatical tagging or word-category disambiguation.

In this paper, the challenges and issues in POS tagging with special reference to Marathi have been presented. The Marathi language is one of the major languages of India. It belongs to the Indo-Aryan Language family with about 71,936,894 users. It is predominantly spoken in the state of Maharashtra in Western India (Chaudhari et. al., 2017). In recent years many research institutions and organizations are involved in developing the lexical resources for Marathi for NLP activities. Marathi Wordnet is one such lexical resource developed at IIT Bombay (Popale and Bhattacharyya, 2017).

The paper is organized as follows: Section 2 introduces POS annotation; section 3 provides information on Marathi annotated corpora; section 4 describes Marathi tag set; section 5 explains tagging approaches, section 6 presents ambiguous behaviors of the Marathi words, section 7 presents a discussion on special cases, and section 8 concludes the paper with future work.

2 Parts-Of-Speech Annotation
In NLP pipeline POS tagging is an important activity which forms the base of various language processing applications. Annotating a text with POS tags is a standard low-level text pre-processing step before moving to higher levels in the pipeline like chunking, dependency parsing, etc. (Bhattacharyya, 2015). Identification of the parts of speech such as nouns, verbs, adjectives, adverbs for each word (token) of the sentence helps in analyzing the role of each word in a sentence (Jurafsky D. et. al., 2016). It represents a token level annotation wherein it assigns a token with POS category.

3 Marathi Annotated Corpora
Aim of POS tagging is to create a large annotated corpora for natural language processing, speech recognition and other related applications. Annotated corpora serve as an important resource in NLP activities. It proves to be a basic building block for constructing statistical models for the automatic processing of natural languages. The significance of large annotated corpora is widely appreciated by researchers and application developers. Various research institutes in India viz., IIT Bombay4, IIIT Hyderabad5, JNU New Delhi6, and other institutes have developed a large corpus of POS tagged data. In Marathi, there is around 100k annotated data developed as a part of Indian Languages Corpora Initiative (ILCI)7 project funded by MeitY, New Delhi. This ILCI corpus consists of four domains viz., Tourism, Health, Agriculture, and Entertainment. This tagged data (Tourism - 25K, Health - 25K, Agriculture - 10K, Entertainment - 10K, General – 30K) is used for various applications like chunking, dependency tree banking, word sense disambiguation, etc. This ILCI

4http://www.cfilt.iitb.ac.in/
5https://ltrc.iit.ac.in/
6https://www.jnu.ac.in/
7http://sanskrit.jnu.ac.in/ilci/index.jsp
8https://meity.gov.in/
annotated data forms a baseline for Marathi POS tagging and is available for download at TDIL portal.9

4 The Marathi POS Tag-Set

The Bureau of Indian Standards (BIS)10 has come up with a standard set of tags for annotating data for Indian languages. This tag-set is prepared for Hindi under the guidance of BIS. The BIS tag-set aims to ensure standardization in the POS tagging across the Indian languages. The tag sets of all Indian languages have been drafted by Dept. of Information Technology, MeitY and presented as Unified POS standard in Indian languages11. Marathi POS tag-set has been prepared at IIT Bombay referring to the standard BIS POS Tag-set, IIIT Hyderabad guideline document (Bharati et al, 2006) and Konkani Tag-set (Vaz et. al., 2012). This Marathi POS Tag-set can be seen in Appendix A.

5 Lexical and Functional POS Tagging: Challenges and Discussions

Lexical POS tagging (Lexical or L approach) deals with tagging of a word at a token level. Functional POS tagging (Functional or F approach) deals with tagging of a word as a syntactic function of a word in a sentence. In other words, a word can have two roles viz., grammatical role (lexical POS w.r.t. a dictionary entry) and functional role (contextual POS)12. For example, in the phrase ‘golf stick’, the POS tag of the word ‘golf’ could be determined as follows:

- Lexically it is a noun as per lexicon.
- Functionally it is an adjective as it is a modifier of succeeding noun.

In the initial stage of ILCI data annotation, POS tagging was conducted using the lexical approach. However, over a while, POS tagging was done using the functional approach only. The reason is that, by using the lexical approach we do a general tagging, i.e., tagging at a surface level or token level and by using the functional approach we do a specific tagging, i.e., tagging at a semantic level. This eases the annotation process of chunking and parsing in the NLP pipeline.

While performing POS annotation, many issues and challenges were encountered, some of which are discussed below. Table 1 lists the occurrences of discussed words in the ILCI corpus.

5.1 Subordinators which act as Adverbs

There are three basic types of adverbs. They are time (N_NST), place (N_NST) and manner (RB). Subordinators are conjunctions which are tagged as CCS.

However, there are some subordinators in Marathi which act as adverbs. For example, वाईटे (tyApramANe, like-wise), ला (tyApramANe, like that), वाईटेल्या (hyAp- RamANe, like this), वाईटेल्या (hyApAmuLe, because of it), वाईटेल्या (hyApAmuLe, because of that), जोगें (jogA, jogyA, when) and लेचा (tevA, then). वाईटेल्या (tyApramANe) and वाईटेल्या (tyApramANe) are generated from pronominal stems viz., वा (yA) and वा (hyA) hence they are lexically qualified as pronouns, however, they function as adverbs; hence to be functionally tagged as RB at the individual level. However, when these words appear as part of the clause then they should be functionally tagged as CCS.

This distinction was also observed by noted Marathi grammarian, Damle13 (Damle, 1965) [p. 206-07].

5.2 Words with Suffixes

There are suffixes like मुळे (muLe, because of; due to), ते (saThI, for), बरोबरे (barobara, along with), etc. When these suffixes are attached to pronouns they function as adverbs or conjunctions at a syntactic level. For example, words लामुळे (tyAmuLe, because of that), लामुळे (tyAmuLe, because of this), लामुळे (hyAmuLe, because of this), लामुळे (hyAmuLe, because of it), लामुळे (hyAmuLe, because of them), यासाठी (yAsAThI, for this), यासाठी (yAsAThI, for that), यासाठी (yAsAThI, at this time), यासाठी (yAsAThI, after this), यासाठी (yAsAThI, after that).

5.3 Words which are Adjectives

Adjectives are tagged as JJ. Consider the example below: जोगें (jogA, doable) is lexically tagged as a verb. However, it should be tagged as PRP. Functionally these words act as conjunctions at the sentence level; therefore, they should be tagged as N_NST. Also, consider the words लामुळे (tyAmuLe, at that time), लामुळे (hyAmuLe, at this time), लामुळे (tyAmuLe, after this), लामुळे (tyAmuLe, after that). Here, wherever the first string/morpheme appears as ला (yA) and वा (hyA), the tag should be given as PRP, lexically. But functionally, all these words shall be tagged as N_NST (time adverb).

5.4 Adnominal Suffixes Attached to Verbs

The adnominal suffix जोगे (jogaM) and all its forms (जोगे, जोगें, जोग, जोगेल्या, jogA, jogI, joge, jogyA) are always attached to verbs. For example, word करणाले (karANa) (karaNyA, doable) is lexically tagged as a verb. However, word करणाले (karANa) is a Kranda form of a verb करणे (karaNe, to do) and suffix जोगे (jogaM) is an adnominal suffix attached to Kranda form; hence, a verb with all the

9 http://www.tdil-dc.in/
10 http://www.bis.gov.in/
11 http://tdil-dc.in/tdildcMain/articles/134692Draft POS Tag standard.pdf
13 http://www.cfilt.iitb.ac.in/damale/index.html
forms of जोगं (jogam) should functionally be treated as adjectives. Therefore verbs with adnominal suffix should be tagged as JJ.

5.5 Words जसें (jase) तसें (tase)

As per Damle (1956), words जसें (jase, like this) and तसें (tase, like that) are tagged as adverbs. However, if they appear with nouns in a sentence, they are influenced by the inflection and gender property of that nominal stem. For example, words जसें (jase, like this) and तसें (tase, like that) have inflected forms like जसा (jasa, like him), जसी (jashi, like her), तसा (tasA, like this), तसी (tashi, like this), तसे (tase, like this), etc. All these words function as a relative pronoun in a sentence. Hence, the words and their variations should be functionally tagged as PRL.

5.6 Word तसेतर (tasetara)

A word तसेतर (tasetara, as it is seen) is the same as तसे पाहिले तर (tase p4hile tara, as it is seen). Lexically, it can be tagged as a particle (RPD) but since it has a function of conjunction; it should be tagged as CCD. For example, in a sentence तसेतर तातकामुळे काढिले वलये वेतात (tasetara taNAdAmaLehI kAlI valaya yetAta), as it is seen that black circles appear because of stress as well), word तसेतर (tasetara) functions as conjunction and hence should be tagged as CCD instead of tagging it as RPD.

5.7 Word अन्यथा (anyatha)

The standard dictionaries give POS of the word अन्यथा (anyatha, otherwise; else; or) as an adverb/indeclinable. For example, consider a sentence अन्यथा तो येणार नाही (anyatha t4 to yeN4ara nAhI, Otherwise, he will not come). Here, while annotating अन्यथा (anyatha) there is a possibility that annotator can directly tag this word as an adverb at a lexical level. However, it behaves like conjunction at the sentence level and hence it should be tagged as CCD.

5.8 Different Forms of कसा (kasA)

As per BIS Tag-set, words कसा, कणी, कळे (kasA, kashi, kase; how) shall be tagged as PRQ. However, the PRQ tag is only for pronom category and the word कसा (kasA) is not a pronoun; it can behave as an adverb or as a modifier. Consider the examples below:

1. तो माणूस कसा आहे हे ल्याच्याचे वैशिष्ट्यांतर कवचल (to m4NUsa kasA Ahe he tyAchAshI bolalyAvaRachA ka4lela, we will come to know about him only after talking to him) [adnominal]
2. सकारात ठरावाचे कायचे कलम कसा रंग होणार (sarakArdhI ThAraVAne kAva4Ache kAlama kase rAndA hoN4ara, How can this clause of law be prohibited by Government Resolution?) [adverbial]

In the 1st case, word कसा (kasA, how) functionally acts as a pronoun, hence to be tagged as PRQ. While, in the 2nd case, it acts as an adverb, hence to be functionally tagged as RB.

5.9 Word मात्र (mAtra)

A word मात्र (mAtra) is very ambiguous in its various usages; it is difficult to functionally identify the POS of this word at a sentence level. Various meanings of word मात्र (mAtra) are given in Data-Karve dictionary14. Some of the different senses of मात्र (mAtra) are discussed here:

- When the word मात्र (mAtra) conveys the meaning of ही, देखील, हीले (hi, dekhila, suddhi; also) then it should be tagged as RB functionally.
- When a word is related to the preceding word तेथे (tethe, there) and its function is an emphatic marker च (cha) then it should be tagged as RPD functionally.
- When word मात्र (mAtra) appears in the form of conjunction then it should be marked as CC functionally.
- If the word is modifying the succeeding noun, then it should be tagged as JJ functionally.
- If the word is modifying the preceding word, then the tag will be RPD as a particle functionally.

Therefore, it is noticed that the word मात्र (mAtra) does not have one single POS tag functionally and it depends upon the appearance in a sentence. Hence, should be tagged as per the usage.

<table>
<thead>
<tr>
<th>Token</th>
<th>Lexical</th>
<th>Functional</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>जण्डमात्र</td>
<td>Pronoun</td>
<td>Adverb</td>
<td>269</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Pronoun</td>
<td>Conjunction</td>
<td>530</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Pronoun</td>
<td>Conjunction</td>
<td>424</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Pronoun</td>
<td>Conjunction</td>
<td>8</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Pronoun</td>
<td>Time adverb</td>
<td>244</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Pronoun</td>
<td>Time adverb</td>
<td>33</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Pronoun</td>
<td>Time adverb</td>
<td>71</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Pronoun</td>
<td>Time adverb</td>
<td>298</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Adjective</td>
<td>Adverb</td>
<td>104</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Adjective</td>
<td>Adverb</td>
<td>246</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Relative pronoun</td>
<td>Adverb</td>
<td>1007</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Relative pronoun</td>
<td>Adverb</td>
<td>511</td>
</tr>
<tr>
<td>जण्डमात्र</td>
<td>Verb</td>
<td>Adjective</td>
<td>97</td>
</tr>
</tbody>
</table>

Table 1: Occurrences of discussed words and lexical v/s functional tags assigned to these words

6 POS Ambiguity: Challenges and Discussions

Ambiguity is a major open problem in NLP. Several POS level ambiguity issues were faced by annotators while annotating the Marathi corpus. Following are some POS

http://www.transliteral.org/dictionary/mr.kosh.maharashtra/source
specific ambiguity problems encountered while annotating.

6.1 Ambiguous POS: Adjective or Noun?
Examples: वयस्कर (vayaskara, the aged)
- कुठंबाच्या वयस्कर सदस्यांनी मतदन केले (kuTuM-bAchya vayaskara sadasyAMni matadAnakele, all the aged members of the family voted).
- त्या वयस्करांनी मतदन केले (svary vayaskarAMni matadAna kele, all the aged people voted).

In the above examples, the word वयस्कर (vayaskaraAMni) lexically acts as an adjective as well as a noun. However, at the syntactic level, in the first example, it is functioning as adjective hence to be tagged as JJ, while in the second example it is functioning as a noun hence to be tagged as N_NN. This is one of the challenges while annotating adjectives appearing in nominal form. Annotators usually fail to disambiguate these types of words at the lexical level; therefore such words should be disambiguated at syntactic level. Hence, annotators need to take special care while annotating such cases.

6.2 Ambiguous POS: Demonstrators
While annotating demonstrators such as हा, ही, हे, ती, ते ((hA, hI, he), this), (to, ti, te), that) annotators often get confused whether to tag them as DMD or DMR. Simple guideline can be followed is, if the demonstrator is directly following noun, then tag it as DMD, otherwise tag it as DMR i.e., if the demonstrator is referring to previous noun/person.

6.3 Ambiguous POS: Noun and Conjunction
Example: word कारण (karaNa, reason; because). At semantic level, the word कारण (karaNa) has two meanings, one is ‘a reason’ which acts as a noun and another is ‘because’ which acts as a conjunction. Annotators have to pay special attention while tagging such cases.

6.4 Ambiguous Words: ते (te) and तेही (tehI)
The word ते (te) has different grammatical categories like pronoun (they), demonstrator (that) and conjunction (to).
Examples:
- ३० ते ४० (30 te 40, 30 to 40)
- The word ते (te) lexically and functionally acts as conjunction, hence to be tagged as CCD.
- ते म्हणाले (te mhanałe, they said)
- Here word ते (te) acts as personal pronoun, hence to be tagged as PR. PRP
- ते कुठे आहे? (te kuThे The Aheta?, where are they?)
- Here word ते (te) acts as relative demonstrator, hence to be tagged as DM_DMR
- रकेशाने गोष्टीसाठी कोणते कारण ते दोन्ही चोर पकडले गेले (rAkeShane pollsAMNa phona keLa ANI te donhI chora pakaDale gele, Rakesh called police and those two thieves got caught).
- Here, word ते (te) is modifying its succeeding noun चोर (chora, thief) so it is Deictic demonstrator, hence to be tagged as DM_DMD.

- लांगे हे करींच पसंत नबते, लांग्धा मुलाने संगीत विकावे आणि तेही नृत्य (tyAMna he kadhIchapasaMnave-hate, tyAMchyaAmuliAmesAMgItashtikAveANitE-hinRitiya, He never wanted his son to learn music and that too the dance form)
Here, the word तेही (tehI) is an ambiguous word. It is modifying succeeding noun or previous context. Here, ही (hI) is a bound morpheme and conveys the meaning ‘also’. Therefore word तेही (tehI) should be tagged as DM_DMR.

6.5 Ambiguous word: उलटा (ulaTA)
Examples:
- उलटे उपायातील कुटुंबांना जाती (ulaTe TAMuna sukavale jAte). Here, उलटे (ulaTe, upside down is behaving as manner, not a noun, hence to be tagged as RB.
- उलटे बांधू सुटून करू (ulaTe bhAMDe sulaTe kara). Here उलटे (ulaTe) it is modifying succeeding noun, hence it is an adjective, hence to be tagged as JJ.

In the above examples, annotator should identify word behavior in the sentence and tag accordingly.

6.6 Ambiguous words: कितीही (kiilhI), ना का (nA kA) and असू दे ना का (asU de nA kA)
Examples:
- संगणक हा कितीही प्रमाण किंवा चर्चा असू दे ना का, तो केवल तेच काम करून शकतो ज्ञानी विविषी (प्रदर्श) आप्स्त्रयला लागू ह्याते. (saMaGnaNaka hA kiilhI pragata kiMvA chatura asU de nA kA, to kevala techa kAma kARa jyACHI viDHi (paddha-ta) ApyAlA sva: maHItA ahe. The computer how much ever may be advanced and clever, it only does that work whose method we only know). Here, कितीही (kiilhI, how much) is a quantifier, hence to be tagged as QTF.
- असू दे ना का (asU de nA kA), the token ना (nA) is a part of verb असू दे (asU de, let it be) and should be tagged as VM, hence the phrase should be tagged as VM, while the token का (kA) is acting as a particle in this phrase and not as a question marker, therefore का (kA) should be tagged as RPD.
- किती हां जेवायला होती? (kiilh maNase jevAyalA hotI, how many people were there for a meal?). Here, किती (kiilh, how many) is a question so it should be tagged as DMQ.

6.7 Ambiguous word: तारा (tara)
Examples:
- Conjunction: कर भी केवळ गेलो नसतो तर हा वाला नसतो (jAra mI veLicha gela nasato tara hA vA-chala nasata, If I had not gone on time he would have not survived).
- Particle: हो! आता भी जातो तर! = ‘भी अबिज्ञात जाणार नाही’ (ho! AIa mI jAto tara! = 'mI ajibAja AINA-
ra nAhI’, ‘yes! now I am leaving then’ = ‘I am not at all leaving’.

In the above sentences, the word तर (tara) is used as a supplementary or stressable word so somewhat special as to give meaning in the sentence. (Date-Karve, 1932). Hence it should be treated as CCD.

- तुम्ही तर लाख केवळ मागलें व मी तर केवळ गरीब पालणी (tumhI tara lakha rupaye mAgatAM va ml tara kevaLa garIba pAdalo, you are asking for lakh rupees and I am a poor person). In this sentence, the word तर (tara) indicates opposition with respect to meaning between two connected sentences. (Date-Karve, 1932). Hence, it should be treated as a RPD.

7 Discussions on Some Special Cases

- Words अमलयुक्त (Amlayukta), मललहित (malalrahitA), मदेदरहित (medarahitA), दुरशक्तुस्त्र (duShkaLgrasta) are combinations of noun plus adjective suffix such as युक्त (yuktA), प्रा (grasta) and रहित (rahitA). In such cases, even though noun is a head string and adjective part is a suffix, the whole word shall be tagged as JJ.
- Before tagging अभंग (abhaMga, verses), ओय्या (ovyA, stanzas), कव्य (kAvya, poetry), etc., annotator shall first read between the lines; understand the meaning which it conveys and then decide upon the grammatical categories of each token. For example, in sentence कौणतेही कार्यान ता क्रले क्षेत्रकाळे कत्रणा ते सावे सावे अंगी (koNatAha karaNa te sache sache sAcha aMgI) the POS tagging should be done as सावेV_VM सावेV_VM अंगीN_NN अंगीN_NN, etc.
- Doubtful cases of word कोणता (koNatA)
  - Examples:
    - कौणता मुलगण हुआर आहे (koNatA mulaGa hu-shAra Ahe)?
    - जाहिकलिच दरम्यान कोणताही हानी झालेली नाही (jaHatuklicChA daramyAna koNatIhI haNi jhAlEli laNhI).
    - ह्या शेताकार्या मांदार निवारण वाहतुकी आहेत (hyaMkhyA boLaNyaVachA mAchyAVara koNatAhI pariNMAna jhAlA nAhI).
    - शेताकार्यस कोणताही व्यऱ्या पाण्याची कमतरता भावाताना भावाची नाही (shetakaryAsa koNatAhI varShI pAnyaVachI kamatarAta bhAsaAnaRra nAhI).

Here, in the 1st example, the word कोणता (koNatA, which one) undoubtedly is DMQ. In rest of the examples कोणताही (koNatAhI, whichever, whomever), कोणताही (koNatAhI, whichever, whomever), कोणताही (koNatAhI, whichever, whomever) are DM adjective (DMD).

8 Conclusion and Future Work

Marathi POS tagging is an important activity for NLP tasks. While tagging, several challenges and issues were encountered. In this paper, Marathi BIS tag-set has been discussed. Lexical and functional tagging approaches were discussed with examples. Further, various challenges, experiences, and special cases have been presented. The issues discussed here will be helpful for annotators, researchers, language learners, etc. of Marathi and other languages.

In future, more issues such as tagging for words having multiple senses; words having multiple functional tags will be discussed. Also, tagset comparison of close languages will be done. Further, the evaluation of lexical and functional tagging using statistical analysis will be done.

References


Edna Vaz, Shantaram V. Walawalkar, Dr. Jyoti Pawar, Dr. Madhavi Sardesai. (2012). BIS Annotation Standards With Reference to Konkani Language. 24th International Conference on Computational Linguistics (COLING 2012), Mumbai.


Yashwant Ramkrishna Date, Chintman Ganesh Karve, Aba Chandorkar, Chintaman Shankar Datar. (1932). Maharasthra Shabdakosh. Published by H. A. Bhave, Varada Books, Senapati Bapat Marg, Pune.
## Appendix A

### Marathi Parts of Speech Tag-Set with Examples

<table>
<thead>
<tr>
<th>Sl. No</th>
<th>Category (क्रियानिर्माणक क्रिया)</th>
<th>Label</th>
<th>Annotation Convention **</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Noun (नाम)</td>
<td>N</td>
<td>N</td>
<td>गावN_NN गोटातN_NN राहते.</td>
</tr>
<tr>
<td>1.1</td>
<td>Common (कालाभिवतक नाम)</td>
<td>NN</td>
<td>N_NN</td>
<td>गावN_NN गोटातN_NN राहते.</td>
</tr>
<tr>
<td>1.2</td>
<td>Proper (कालाभिवतक नाम)</td>
<td>NNP</td>
<td>N_NNP</td>
<td>राहतेN_NNP गोटातN_NNP राहते.</td>
</tr>
<tr>
<td>1.3</td>
<td>Nloc (स्थल-कार)</td>
<td>NST</td>
<td>N_NST</td>
<td>1. लोयेN_NST काम करते होते. 2. माहीN_NST हातिN_NST हायती आहे.</td>
</tr>
<tr>
<td>2</td>
<td>Pronoun (माही)</td>
<td>PR</td>
<td>PR</td>
<td>मी PR PR म्हणून.</td>
</tr>
<tr>
<td>2.1</td>
<td>Personal (स्वतंत्र वाचक)</td>
<td>PRP</td>
<td>PR_PRP</td>
<td>मी PR PR म्हणून.</td>
</tr>
<tr>
<td>2.2</td>
<td>Reflexive (आत्मवाचक)</td>
<td>PRF</td>
<td>PR_PRF</td>
<td>मी PRF PRF आते.</td>
</tr>
<tr>
<td>2.3</td>
<td>Relative (संबंधी)</td>
<td>PRL</td>
<td>PR_PRL</td>
<td>ज्यानेPR_PRL हे सांगिते त्याने हे काम केले पाहिजे.</td>
</tr>
<tr>
<td>2.4</td>
<td>Reciprocal (परस्परिक)</td>
<td>PRC</td>
<td>PR_PRC</td>
<td>परस्पर</td>
</tr>
<tr>
<td>2.5</td>
<td>Wh-word (भूमिका)</td>
<td>PRQ</td>
<td>PR_PQi</td>
<td>कोणाचेPR_PQi कोणती.PR_PQi हाय. कोणाचेPR_PQi हाय. PR_PQi आहे.</td>
</tr>
<tr>
<td>2.6</td>
<td>Indefinite (अनिर्णित)</td>
<td>PRI</td>
<td>PR_PRI</td>
<td>कोणाचेPRI कोणतीPR_PRI हाय. कोणाचेPRI कोणतीPRI हाय. कोणाचेPRI कोणतीPRI हाय.</td>
</tr>
<tr>
<td>3</td>
<td>Demonstrative (दर्शक)</td>
<td>DM</td>
<td>DM</td>
<td>हे DM DM माही आहे.</td>
</tr>
<tr>
<td>3.1</td>
<td>Deictic</td>
<td>DMD</td>
<td>DM_DMD</td>
<td>ती DM DM माही आहे.</td>
</tr>
<tr>
<td>3.2</td>
<td>Relative</td>
<td>DMR</td>
<td>DM_DMR</td>
<td>हे DM DM माही आहे.</td>
</tr>
<tr>
<td>3.3</td>
<td>Wh-word</td>
<td>DMQ</td>
<td>DM_DMQ</td>
<td>कोणता DM DM माही आहे?</td>
</tr>
<tr>
<td>4</td>
<td>Verb (क्रियापद)</td>
<td>V</td>
<td>V</td>
<td>बसा V V.</td>
</tr>
<tr>
<td>4.1</td>
<td>Main (मुख्य क्रियापद)</td>
<td>VM</td>
<td>V_VM</td>
<td>तो परी VM.</td>
</tr>
<tr>
<td>4.2</td>
<td>Auxiliary (सहायक क्रियापद)</td>
<td>VAUX</td>
<td>V_VAUX</td>
<td>राज परी VM.</td>
</tr>
<tr>
<td>5</td>
<td>Adjective (विश्वसनीय)</td>
<td>JJ</td>
<td>JJ</td>
<td>तुबद्ध JJ माही.</td>
</tr>
<tr>
<td>6</td>
<td>Adverb (क्रियाविवेच)</td>
<td>RB</td>
<td>RB</td>
<td>हा RB.</td>
</tr>
<tr>
<td>7</td>
<td>Conjunction (उपयुक्त विरुध्ध)</td>
<td>CC</td>
<td>CC</td>
<td>तावा CC CC भांत.</td>
</tr>
<tr>
<td>7.1</td>
<td>Coordinator</td>
<td>CCD</td>
<td>CC_CCD</td>
<td>तावा CC_CCD भांत.</td>
</tr>
<tr>
<td>7.2</td>
<td>Subordinator</td>
<td>CCS</td>
<td>CC_CCS</td>
<td>तावा CC_CCS भांत.</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Quotative</td>
<td>UT</td>
<td>CC_CCS_UT</td>
<td>तावा CC_CCS UT भांत.</td>
</tr>
<tr>
<td>8</td>
<td>Particle</td>
<td>RP</td>
<td>RP</td>
<td>नो RP RP वसा.</td>
</tr>
<tr>
<td>8.1</td>
<td>Default</td>
<td>RPD</td>
<td>RP_RPD</td>
<td>मो RP RPD वसा.</td>
</tr>
<tr>
<td>8.2</td>
<td>Intercor (उपर वाचक)</td>
<td>INJ</td>
<td>RP_INJ</td>
<td>अरे RP_INJ.</td>
</tr>
<tr>
<td>8.3</td>
<td>Intensifier (ढंढ क्रियापद)</td>
<td>INTF</td>
<td>RP_INTF</td>
<td>तर्क RP_INTF.</td>
</tr>
<tr>
<td>8.4</td>
<td>Negation (नकारक)</td>
<td>NEG</td>
<td>RP_NEG</td>
<td>नकार, न</td>
</tr>
<tr>
<td>9</td>
<td>Quantifiers</td>
<td>QT</td>
<td>QT</td>
<td>पहेळात QT QT.</td>
</tr>
<tr>
<td>9.1</td>
<td>General</td>
<td>QTF</td>
<td>QT_QTF</td>
<td>पहेळात QT QTF.</td>
</tr>
<tr>
<td>9.2</td>
<td>Cardinals</td>
<td>QTC</td>
<td>QT_QTC</td>
<td>माही QTC QTC भांत.</td>
</tr>
<tr>
<td>9.3</td>
<td>Ordinals</td>
<td>QTO</td>
<td>QT_QTO</td>
<td>माही QTO QTO.</td>
</tr>
<tr>
<td>10</td>
<td>Residuals (अनीसिव)</td>
<td>RDF</td>
<td>RD_RDF</td>
<td>अनीसिव RD_RDF.</td>
</tr>
<tr>
<td>10.1</td>
<td>Foreign word</td>
<td>RD</td>
<td>RD</td>
<td>अनीसिव RD.</td>
</tr>
<tr>
<td>10.2</td>
<td>Symbol</td>
<td>SYM</td>
<td>RD_SYE</td>
<td>$, &amp;,* (,).</td>
</tr>
<tr>
<td>10.3</td>
<td>Punctuation</td>
<td>PUNCT</td>
<td>RD_PUNC</td>
<td>: (period), , (comma), , (semi-colon), ? (exclamation), ? (question), : (colon), etc.</td>
</tr>
<tr>
<td>10.4</td>
<td>Unknown</td>
<td>UNK</td>
<td>RD_UNK</td>
<td>Not able to identify the Tag.</td>
</tr>
<tr>
<td>10.5</td>
<td>Echo-words</td>
<td>ECH</td>
<td>RD_ECH</td>
<td>जेवण ECH.</td>
</tr>
</tbody>
</table>