

Morphological Generator for Telugu Nouns and Pronouns

Sasi Raja Sekhar Dokkara
Dept. of Computer Science
Adikavi Nannaya University,
India

Suresh Varma Penumathsa
Dept. of Computer Science
Adikavi Nannaya University,
India

Somayajulu G. Sripada
Dept. of Computing Science
University of Aberdeen, UK

ABSTRACT

Telugu is a free word order language in which various grammatical categories (case, gender, number, person etc.) are morphologically encoded making it a morphologically rich language. In this paper we present a morphological generator for Telugu nouns and pronouns modelled on finite state techniques. The morphological generator generates the required word form for nouns and pronouns from an input specification consisting of the lemma and its associated features. The current work is an independent module of a surface realization engine that automates the task of building grammatically correct Telugu sentences. The morphological generator also supports verb morphological generator in generating the appropriate verb form by passing the appropriate features (person, number and gender) required for the formation of the appropriate verb form.

General Terms

Natural Language Processing (NLP), Morphological Analyser (MA), Plural Formation, Oblique Stem Formation, Case Marker agglutination.

Keywords

Natural Language Generator (NLG), Morphological Generator (MG), Finite Automata, Nouns and Pronouns, Number, Gender, Person.

1. INTRODUCTION

Natural Language Processing (NLP) applications which are growing in number these days can be categorized into two broad areas namely Natural Language Understanding (NLU) where Morphological Analysers (MA) play a very important role and Natural Language Generation (NLG) where Morphological Generators play a very important role. A Morphological Analyser takes a word as input, processes it and gives as output its root along with its grammatical features. A Morphological Generator takes a root along with its grammatical features as input and generates the required word form. Morphological Generators (MG) have a very important role to play in applications like Surface Realization (Dokkara, Penumathsa, and Sripada, 2015, Gatt and Reiter, 2009) and Machine Translation (Kristina Toutanova, Hisami Suzuki, Achim Ruopp, 2008) of free word order languages like Telugu. It is always advantageous for free word order languages like Telugu to have Morphological Generator as a separate component that is separate from the rest of the NLG system (Guido Minnen et.al 2000). The current work is a separate module of a surface realization engine for Telugu (Dokkara, Penumathsa, and Sripada, 2015), a java application which is the final subtask of a Natural Language Generation (NLG) pipeline (Reiter and Dale, 2000). The sentence realization engine for Telugu is designed following the

SimpleNLG (Gatt and Reiter, 2009) approach which is a very popular surface realization engine for English.

Telugu is a morphologically rich free word order language spoken by people from the south Indian states of Andhra Pradesh and Telangana. In this paper we describe a morphology engine which automatically generates the different forms of nouns and pronouns in Telugu. The current work is modelled on the morphology engine for English (Guido Minnen et.al 2000).

Telugu nouns are divided into three classes namely (i) proper nouns and common nouns, (ii) pronouns, and (iii) special types of nouns (e.g. numerals) (Krishnamurti and Gwynn, 1985). All nouns except few special type nouns have gender, number, and person. Noun morphology involves mainly plural formation, oblique stem formation and case inflection. The current work discusses in detail the first two classes of nouns.

2. RELATED WORK

Linguistic theories of morphology differ in their view with respect to treating morpheme as the basic building block of morphological analysis or generation. "Two models of grammatical description" was proposed by (Hockett, 1954) which are Item and Arrangement model (IA) and Item and Process model (IP). Item and Arrangement model is a concatenative approach where morphemes are the lexical units and morphology is an agglutination of such units to form words. Item and Process (IP) is considered as a derivational process where new word forms can be produced by acting on morphemes and words. He also mentioned about an already existing model the Word and Paradigm (WP) which is a word based morphological approach which states generalizations that hold between the different forms of inflectional paradigms and used in languages like Latin, Greek, and Sanskrit. A two dimensional taxonomy of morphological theories was proposed by (Stump, 2001). He distinguished two axes along which inflectional morphology may be situated relative to one another. He proposed the lexical/inferential axis and the incremental/realizational axis (Stump, 2001) which are orthogonal to each other.

From a computational perspective lexical-incremental and inferential-realizational are computationally equivalent and can be implemented using finite state techniques. A diverse range of languages have used finite state techniques to build morphological analysers (MA) and morphological generators (MG) (Kartunen, 2003, Beesley and Kartunnen, 2003, Kartunen and Beesley, 2005, Roark and Sproat, 2007). Therefore, in this paper we apply finite state techniques to Telugu morphology of nouns and pronouns.

Among the morphological tools for Indian languages (Goyal and Lehal 2008) report a machine translation system from Hindi-Punjabi modelled on database approach where all word forms are stored in relational database. A number of morphological tools for Tamil are reported by (Antony and Soman 2012) which range from corpus based through suffix stripping to finite state techniques. For Telugu language, (Rao et.al 2006) describe a word and paradigm based morphological analyser and generator. (Sribadrinarayan et al 2009) describe an item and arrangement based morphological generator for Telugu. (Ganapatiraju et al 2006) describe a rule based (item and process based) morphological generator for Telugu.

3. INPUT SPECIFICATION

The noun and pronoun morphology engine presented in this paper is an independent module of a surface realization engine which produces grammatically correct Telugu sentences. The input specification for the surface realization engine consists of lexicalized grammatical constituents and associated features in the form of an XML file. The input specification provides the required grammatical information not only at the sentence level but also at the word level. An example XML specification corresponding to the Telugu sentence (1) is shown in Figure 1.

```
vAlYIYu iMtiki vaccAru. (They came home.) (1)
<?xml version="1.0" encoding="UTF8" standalone="no">
<document>
<sentence type="" predicatetype="verbal" respect="no">
<nounphrase role="subject">
<head pos="pronoun" gender="human" number="plural"
person="third" casemarker="" stem="basic">vAdu</head>
</nounphrase>
<nounphrase role="complement">
<head pos="noun" gender="nonmasculine"
number="singular" person="third" casemarker="ki"
stem="oblique">illu</head>
</nounphrase>
<verbphrase type="">
<head pos="verb" tensemode="pasttense">vaccu</head>
</verbphrase>
</sentence>
</document>
```

Figure1. XML Input Specification

A very popular transliteration scheme for representing Indian language alphabets in ASCII character set called as WX notation (see appendix B in Bharati et al, 1995) is used to specify Telugu words in Figure 1. In WX notation the small case letters are used for un-aspirated consonants and short vowels while the capital case letters are used for aspirated consonants and long vowels. The retroflexed voiced and voiceless consonants are mapped to 't, T, d and D'. The dentals are mapped to 'w, W, x and X'. Hence the name of the scheme "WX", referring to the idiosyncratic mapping. This scheme is widely used in Natural Language processing in India and a lot of tools are available for conversion between WX notation and Indian languages. We independently developed a tool which converts the lexicalized input given in WX notation to Telugu sentence as output. This tool is used in the surface realization engine for Telugu (Dokkara, Penumathsa, and Sripada, 2015).

4. NOUN MORPHOLOGICAL PROCESS

The head of the complement in the example (1) has one noun "iMtiki". The word "illu" given in the XML specification of Figure 1 as the head of the noun phrase which plays the role of a complement in the sentence along with its feature values can be written as follows:

"illu", noun, nonmasculine, singular, third, oblique, "ki"---iMtiki

First the oblique stem of the word "illu" is formed as the word needs to get agglutinated with the case marker. The formation of the oblique stem is a two-step process. In the first step the class to which the root word belongs is identified. In the current work the identification of the class is modelled on finite state techniques. The root word "illu" belongs to class III. A pictorial representation of the finite automata for identification of class III stems is shown in Figure 2. In the second step the oblique stem of "illu" which is "iMti" is formed by replacing "llu" by "Mti" (see sec. 5.2.1 for details).

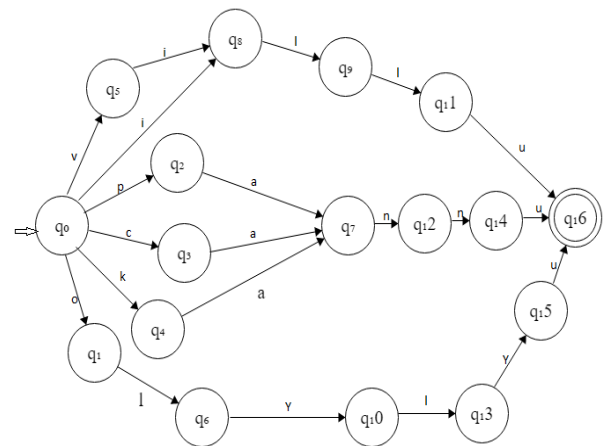


Figure 2: A finite automata to identify class III stems for oblique stem formation

After the formation of the oblique stem the case marker gets agglutinated to the oblique stem to form the final word as follows:

"iMti" + "ki"----- "iMtiki".

The formation of the pronoun "vAlYIYu" in the example of (1) which is the head of the subject along with its feature values can be written as follows:

"vAdu", pronoun, human, plural, third, basic, ""-"vAlYIYu"

The formation of plurals for pronouns does not have any rules and therefore they are stored in a lexicon. The word "vAlYIYu" is retrieved from the lexicon stored in an external file "pronounplural.xml" as the plural for the word "vAdu" and the feature values.

5. MORPHOLOGY ENGINE

The steps taken by the noun or pronoun root to get the required inflection form are as follows:

- Formation of the plural of the root if required.
- Formation of the oblique stem if required.
- Agglutination of the case-marker to the oblique stem if required.

5.1 Plural Formation

Common nouns can be divided into count and non-count nouns. Non-count nouns (mass nouns, indivisible objects and abstract nouns) cannot be distinguished for number they are either singular or plural. Some count nouns do not conform to any rules of plural formation. The singular and plural forms of the non-count nouns and the count nouns which do not conform to any rules of plural formation are stored externally in a lexicon as an XML file named “plural.xml”.

Some mass nouns that exist only in the singular form are given in Table 1.

Table 1: Example Mass nouns in singular

Word	Meaning in English
uppu	Salt
nUne	Oil
inumu	Iron
veVMdi	Silver
biyyaM	Rice
janaM	People

Some mass nouns that exist only in the plural form are given in Table 2.

Table 2: Example Mass nouns in plural

Word	Meaning in English
vadlu	Paddy
peVsalu	Green gram
kaMxulu	Red gram
nIlyIYu	Water
pAlu	Milk

Indivisible objects cannot have both singular and plural forms. Some indivisible objects are shown in Table 3.

Table 3: Example Indivisible objects

Word	Meaning in English
AkAsaM	Sky
samuxraM	Sea

Some example abstract nouns which are non-count nouns are shown in Table 4.

Table 4: Example Abstract nouns

Word	Meaning in English
welupu	Whiteness
welivi	Intelligence
balaM	Strength
saMwoRaM	Happiness
nixra	Sleep

Among the count nouns some nouns do not have any rules for the formation of plurals. Table 5 is a list of count nouns which do not conform to any rules for the formation of plurals:

Table 5: Count nouns which do not confirm to any rules

Singular Word	Plural Word
rAyi	rAIYIYu
Poyyi	poyyilu
peMdli	peMdliMdlu
Vari	vadlu
gAru	gArlu
sAri	sArlu
kumArudu	kumArulu
eVxxu	eVdlu
veVyyi	velu
Cenu	celu
Penu	pelu
kAdi	kAMdlu
edu	eMdlu
jIwagAdu	jIwagAIYIYu
alludu	allulIYIYIYu
manamarAlu	manamarAIYIYu
ceVlleVlu	ceVlleVIYIYu
kUwuru	kUwulYIYu
koVdavali	koVdavalYIYu
rAwri	rAwriIYIYu
bAvi	bAvulu
paMci	paMculu
pilli	pillulu
puli	pululu
saMgawi	saMgawulu
maniRi	manuRulu
kolimi	kolumulu
muliki	mulukulu
eVnimixi	eVnumuxulu
awiXi	awiXulu
pariXi	pariXulu
samiwi	samiwulu

Pronouns do not conform to any rules regarding the formation of plurals. All the pronouns and their plurals listed in Table 6 are also stored externally in a lexicon as an XML file named “pronounplural.xml”.

Table 6: Example pronouns and their plurals

Person	Singular	Plural
First	nenu	memu
Second	nuvvu	mIru
Third	vAdu	vAIYIYu
	axi	avi
	vAru	vAIYIYu
	ixi	ivi
	awanu	vAIYIYu
	Ayana	vAIYIYu
	AmeV	vAIYIYu
	imeV	vIlyIYu
	Avida	vAIYIYu
	vIdu	vIlyIYu
	iwanu	vIlyIYu
	Iyana	vIlyIYu
	vIru	vIlyIYu
	Ivida	vIlyIYu
	exi	evi
wAnu	wAmu	

The regular way of forming the nominative plural is by adding the plural suffix “lu” to the basic stem.

Example:

Avu (cow)→Avulu (cows)

A number of morphophonemic changes may occur because of which sometimes the plural suffix “lu” become “IYIYu”.The morphophonemic changes occur based on the class to which the singular stem belongs. The formation of the plural is a two-step process. First the class to which the stem belongs is identified. In the current work the identification of the class to which the stem belongs is implemented as a finite automata illustrated in Figure 3 for class VII. Second the sandhi (morphophonemic changes) occurs and the final plural form is generated.

The stems can be categorized into different classes for plural formation of nouns (Krishnamurti and Gwynn, 1985) as follows:

- I) Stem final ending in “i/u” preceded by “t”, “Mt”, “Md” is lost before the plural suffix “lu”

Example:

koti (one crore)→kotlu (crores)

- II) In all stems ending in “di”, “du”, “lu” and “ru” and in stems of more than two syllables ending in “li” and “ri” the final syllable becomes IYIY before IYIYu

Example:

badi (school) →baIYIYu (schools)

Exception 1:

Masculine nouns of Sanskrit origin ending in “du” replace “du” by “lu” to form the plural

Example:

snehiwudu (friend) →snehiwulu (friends)

Exception 2:

Loanwords from foreign languages ending in “ru” form the plural by adding “lu” to the basic stem.

Example:

nOkaru (servant)→nOkarlu (servants)

- III) Stem final “tti”, “ttu”, “ddi”, “ddu” becomes “t”, “d” before “lu”

Example:

ceVttu (tree)→ceVtlu (trees)

- IV) Stem final following “llu”, “nnu” following a short vowel becomes “Md” or IY before IYu

Example:

illu (house) →ilIYIYu (houses)

Exception 1:

Some stems ending in “nnu” form the plural by adding lu to the basic stem.

Example:

pannu (tax)→pannulu (taxes)

- V) Stem final aM, AM is replaced by A and stem final ending in eVo is replaced by E before the plural suffix lu

Example:

puswakaM (book)→puswakAlu (books)

- VI) Stems ending in Ayi form the plural in the

regular way by adding lu.

Example:

abbAyi (boy)→abbAyilu (boys)

- VII) Stem final ending in “yi”, “yyi” is replaced by “wu” before “lu”. The vowel before “wu” is a long vowel.

Example:

ceVyyi (hand)→cewulu (hands)

The class identification for stems belonging to class VII can be done through the finite automata in Figure 3.

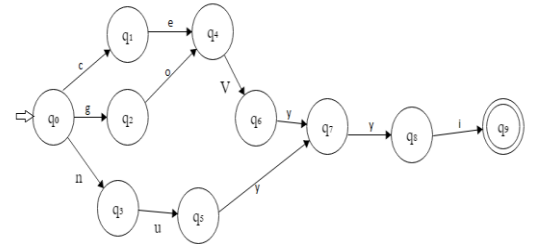


Figure 3: A Finite Automata to identify class VII stems for plural formation

- VIII) Stems that do not confirm to the above classes and when the stem ends in “i”:

- a) If the stem consists of two syllables, or it consists of more than two syllables and the vowel in the middle syllable is not “i” then the final “i” changes to “u” before “lu”.

Example:

bAyi (water well) →bAvulu (water wells)

- b) If the stem consists of two or more syllables and the vowel in the middle syllable is “i” then all the non-initial “i”’s become “u”.

Example:

maniRi (human being)→ manuRulu (human beings)

Exception:

In the nouns of Sanskrit origin the “i” in the middle syllable does not change.

Example: pariXi (boundary)→pariXulu (boundaries)

Proper nouns are not generally used in the plural but when they are used the rules are similar to those of the common nouns.

5.2 Oblique Stem Formation

Each noun in Telugu has an oblique stem along with the basic stem in both singular and plural forms. The oblique stem is used to indicate possession or adjectival relationship. It corresponds in meaning to the possessive forms ‘s (singular), s’(plural) in English.

5.2.1 Oblique Stem in Singular

The oblique stems of the personal pronouns and a few demonstrative pronouns in the singular form like “axi”, “ixi”, “exi” do not confirm to any rules for oblique stem formation. They need to be memorized and are listed in Table 7.

Table 7: pronouns and their oblique stems in singular

Singular Nominative	Singular Oblique
nenu	nA
nuvvu	nI
axi	xAni
ixi	xIni
exi	xeni

The oblique stems in the singular for common nouns and demonstrative pronouns are formed based on some morphophonemic rules. Common nouns in Telugu are divided into six classes based on the manner in which the oblique stem is formed (Krishnamurti and Gwynn, 1985).

The six classes are as follows:

- I) All nouns denoting human beings, demonstrative pronouns ending in “du”, “ru”, “nu”, “lu” and a few non-human nouns ending in “ru”, “lu” preceded by a long vowel fall into this category. These form the oblique stem by deleting the final “u” and adding “i” to the basic stem. Some example nouns and pronouns belonging to Class I are listed in Table 8.

Table 8: Examples for Class I

Singular Nominative	Singular Oblique
mogudu (husband)	mogudi (of husband)
vAdu (he)	vAdi (his)
kuwuru (daughter)	kuwuri (of daughter)
cellelu (sister)	celleli (of sister)
awanu (he)	awani (his)
vAru (they)	vAri (theirs)
kAlu (leg)	kAli (of leg)
Uru (village)	Uri (of village)

- II) Non-human nouns of two or more syllables ending in “du”, “di”, “ru”, “ri”, “lu”, “li” replace the final syllable by “ti” in forming the oblique stem. Some nouns belonging to Class II are listed in Table 9.

Table 9: Examples for Class II

Singular Nominative	Singular Oblique
gUdu (house)	gUti (of a house)
eru (canal)	eti (of a canal)
wAbelu (tortoise)	wAbeti (of a tortoise)
nAgali (plough)	nAgati (of a plough)
kAvadi (balance)	kAvati (of a balance)

- III) Six stems ending in nnu, llu, IYIYu replace them by Mti in forming the oblique stem. All the six stems belonging to Class III are listed in Table 10.

Table 10: Examples for Class III

Singular Nominative	Singular Oblique
illu (house)	iMti (of a house)
villu (bow)	viMti (of a bow)
pannu (tooth)	paMti (of a tooth)

kannu (eye)	kaMti (of an eye)
cannu (breast)	caMti (of breast)
olYIYu (body)	oMti (of body)

- IV) Five stems of two syllables ending in “yi” and two stems ending in “rru” replace the final syllable by “wi” in the formation of the oblique stem. All the seven stems belonging to Class IV are listed in Table 11.

Table 11: Examples for Class IV

Singular Nominative	Singular Oblique
ceyi/ceVyyi (hand)	cewi (of hand)
neyi/neVyyi (ghee)	newi (of ghee)
nUyi/nuyyi (well)	nUwi (of well)
goyi/goVyyi (pit)	gowi (of pit)
rAyi(stone)	rAwi (of stone)
goVrru (harrow)	goVrwi(of harrow)
parru (name ending)	(parwi)

- V) All nouns ending in “M” have two oblique stems, one in the genitive with no modification and the other before the accusative and dative case which is formed by replacing “M” with “Ani”. Some example stems belonging to Class V are listed in Table 12.

Table 12: Examples for Class V

Singular Nominative	Singular Oblique(genitive)	Singular Accusative/Dative
kalaM (pen)	kalaM (pen)	kalAni-
puswakaM (book)	puswakaM (book)	puswakAni-

- VI) Basic stems ending in “e”, “a” or a long vowel, those ending in “i” or “u” preceded by double consonants except “ll” or “nn” and all nouns not covered by classes I-V have both their basic stem and oblique stem identical. Some example stems belonging to Class VI are listed in Table 13.

Table13: Examples for Class VI

Singular Nominative	Singular Oblique
anna (elder brother)	anna (of an elder brother)
peVtteV (box)	peVtteV (of a box)
poti (contest)	poti (of a contest)
ceVttu (tree)	ceVttu (of a tree)

5.2.2 Oblique stem in plural

Telugu language has two words for the English word “we”, one exclusive “memu” which does not include the person who is addressed and one inclusive “manamu” which includes the person who is addressed. The list of personal pronouns and a few demonstrative pronouns like “avi”, “ivi”, “evi” in the plural form do not confirm to any rules for oblique stem formation. They need to be memorized and are listed in Table 14.

Table 14: pronouns and their oblique stems in the plural

Plural Nominative	Plural Oblique
memu/manamu	mA/mana
mfru	mI
avi	vAti
ivi	vIti
evi	veti

The plural oblique stem of the common noun and demonstrative pronoun is formed by uniformly changing “lu” or “IYIYu” to “la” or “IYIYa”. The oblique suffix is “a” added to the plural stem. In Sandhi the final “u” of the plural stem is lost before “a”.

5.2.3 Oblique Stem of Proper Nouns

The oblique stem of proper nouns both singular and plural is formed in the same way as those of common nouns. Table 15 lists some example proper nouns and their oblique stems.

Table 15: Examples for proper noun oblique stems

Proper Noun	Oblique Stem
rAmudu (Rama)	rAmudi (Rama’s)
subbArAvu (Subba Rao)	subbArAvu(Subba Rao’s)
AMXrulu (people from Andhra)	AMXrula (of the people from Andhra)

5.3 Case marker Agglutination

In Indian Languages postpositions (case markers) serve the purpose of prepositions in English. Postpositions which express spatial or temporal relations or mark various semantic roles establish some grammatical relations between the nouns which they follow and the verbs of the sentence. In Telugu postpositions are added to the oblique stems in both singular and plural forms.

Postpositions in Telugu can be classified into two types namely Type-1 and Type-2. Postpositions belonging to Type-1 only occur bound to oblique stems. They never occur as separate words in a sentence or in combination with other postpositions. Most commonly used postpositions of this type are listed in Table 16.

Table 16: List of Type-1 postpositions

Postposition	Meaning
ni/nu	Accusative
ki/ku	Dative
kosaM	for the sake of
Wo	with, along with
nuMci/niMci	from
a/na/ni	in/on/at
kaMteV	than, compared
guMdA/xwArA	Through

Among the Type-1 postpositions dative and accusative can be grouped into a subclass because of the similarity in the morphophonemic changes they exhibit different from the other Type-1 postpositions.

The accusative and dative postpositions, “nu” and “ku” respectively take the form “ni” and “ki” if the preceding syllables end in “i” or “I”, except in the case of personal

pronouns like “nI-ku” (for you), “mI-ku” (for you in plural) with single syllable.

Example:

rAmudu(nominative)+ku → rAmudi(oblique)+ku
→rAmudiki

The use of accusative suffix for nouns denoting inanimate objects is optional. The nouns take the same form as nominative in the accusative.

Example:

amma mAku kaXa ceppiMxi (Mother told us a story)

In the above sentence “kaXa” (story) is an inanimate noun which must have taken the form” kaXanu” but the accusative suffix being optional it takes the nominative form “kaXa” as discussed above.

In the singular of nouns ending in “aM”, “AM”, and “eVM” the dative suffix “ki” and the accusative suffix “ni” are added to variant forms of the oblique stems. The stems ending in “aM” or “AM” (see section 5.2.1 class V) have “Ani” as the variant and stems ending in “eVM” have “eni” as the variant of the oblique stem.

Example:

gurraM + ki →gurrAni-ki (to the horse)
palYIYeVM + ki →palYIYeni-ki (to the plate)

Example sentences of other Type-1 postpositions are as follows:

awanu maxrAsu nuMci vaccAdu (He came from Madras)
rAXa jyowikosam vacciMxi (Radha came for Jyothi)

Postpositions belonging to Type-2 are separate words generally denoting place and time. Although they sometimes occur as postpositions they also occur as independent words mostly as adverbial nouns. A feature of Type-2 postpositions is that postpositions of Type-1 can be added to them for example “lo” which is a Type-2 postposition can be added to “nuMci” and “ki” to form “lonuMci” and “loki”. Most common postpositions of this type are listed in Table 17.

Table 17: List of Type-2 postpositions

Postposition	Meaning
lo	in
lopala	inside
mixa	on
kiMxa	under
bEta	outside
xaggara	near
veVnaka	behind
muMxu	in front of, before
lA,lAgu,lAgA	like
prakAraM	according to
warvAwa	after
varaku,xAkA	up to, until
eVxuta	opposite
maXyana	between
pakkana	by the side of
pAtu	for (of time)
vEpu	in the direction of, towards

Example sentences of Type-2 postpositions are as follows:

mA illu narasApuramlo uMxi (Our house is in Narsapuram)
rAju lopala unnAdu (Raju is inside)
puswakaM ballamixa uMxi (The book is on the table)

6. GENDERS IN TELUGU

In Telugu the gender of nouns and pronouns depend on the number. There are two genders masculine and non-masculine when the number is singular. All nouns denoting male persons belong to the masculine gender and all the others belong to non-masculine gender. There is no feminine gender and all nouns denoting female persons are treated as non-masculine when the number is singular. There are two genders human and non-human when the number is plural. All nouns denoting male and female persons belong to the human gender and all others belong to the non-human gender. The relationship between gender and number is represented in Table 18 (Dokkara, Penumathsa, and Sripada, 2015).

Table 18: Relationship between numbers and genders

Number	Gender
Singular	Masculine, non-masculine
Plural	Human, non-human

As a result two demonstrative pronouns “axi”(that thing , that lady) and “ixi”(this thing, this lady) are non-masculine when the number is singular but when the number is plural they fall into two genders human “vAIYIYu”(those people) and “vIIYIYu” (these people), non-human “avi”(those things) and “ivi”(these things). In the current work both “axi” and “ixi” are treated as referring to things and not female persons because using these words to refer to female persons happens only in casual talk.

Nouns generally do not have any marker of gender but some words and suffixes are used to differentiate between male and female sexes. The different nouns that use suffixes to differentiate between male and female sexes are as follows:

- a) Some masculine nouns end with “rAlu” to indicate the female sex.

Example:
snehiwudu(male friend)
snehiwurAlu(female friend)

- b) Some descriptive words use the suffixes “amma”, “kawwe” to denote female persons and “ayya”, “kAdu” to denote male persons.

Example:
paMwulu (school master)
paMwulamma (school mistress)
musalayya (old man)
musalamma (old woman)
AtakAdu (male player)
Atakawwe (female player)

- c) The word “moVga” (male) and “Ada” (female) are used to distinguish sex in both human beings and animals.

Example:
moVgapilla
Adapilla

- d) Various words are used to distinguish male and female animals and birds.

Example:
kodipuMju (cock)
kodipeVtta (hen)

- e) Among pronouns and numerals certain forms are used to distinguish male and female persons.

Example:
vAdu (he)
AmeV (she)
okkadu (one man)
okawe (one woman)

7. GRAMMATICAL PERSONS IN TELUGU

There are three grammatical persons namely First Person, Second Person, and Third Person in Telugu. They are used to distinguish between the speaker, the addressee, and others. The personal pronouns of Telugu language are defined by the grammatical person.

Verbs in Telugu take a form dependent on the person and number of the subject. Table 19 is a list of verb endings depending on the number and person of the noun or pronoun.

Table 19: Verb endings based on number and person

	Number	Person	Verb Ending
Pronouns and Nouns ending for the verb	Singular	First Person	-nu
		Second Person	-vu
		Third Person	-du/-xi
	Plural	First Person	-mu
		Second Person	-ru
		Third Person	-ru/yi

Example sentences of different grammatical persons with their verb endings are as follows:

nenu annaM winnaA-nu (First Person singular)
nuvvu annaM winnaA-vu (Second Person singular)
vAdu annaM winnaA-du (Third Person Masculine singular)
Ame annaM winna-xi (Third Person Non-Masculine singular)
memu annaM winnaA-mu (First Person plural)
mIru annaM winnaA-ru (Second Person plural)
vAIYIYu annaM winnaA-ru (Third Person Human plural)
avi annaM winnaA-yi (Third Person Non-Human plural)

8. EVALUATION

An evaluation of the performance of the Morphology engine for Nouns and Pronouns is reported here. The evaluation of the morphology engine is performed with respect to the Telugu Noun database downloaded from the web resource [Wiktionary](https://en.wiktionary.org/wiki/Category:Telugu_nouns) at https://en.wiktionary.org/wiki/Category:Telugu_nouns. A total of 524 nouns were downloaded to perform the evaluation. The nouns were tested for plural generation and oblique stem generation and case marker agglutination. Evaluation was not done for pronouns as both plural formation and oblique stems formation for pronouns are

shown in Table No 6 and Table No 7 in section 5.1 and 5.2.1 respectively.

Among the 524 nouns that were downloaded some nouns which were repeated and some of them like “kri . pU” (Before Christ) which are not suitable for pluralization and oblique form generation were removed from the list. A total of 480 nouns were identified to be suitable for the generation of plurals and oblique forms.

The evaluation was performed by giving the nouns as input to the surface realization engine because it speeds up the evaluation as more number of nouns can be tested at the same time.

The evaluation results for the plural formation of the nouns are given in Table 20.

Table 20 : Evaluation Results of Plural Formation

Class	No. of nouns identified in each class
I	6
II	63
III	1
IV	2
V	30
VI	0
VII	0
VIII-a	28
VIII-b	1
Regular	349

The evaluation results in Table 20 show that no nouns are categorized under Class VI and Class VII. Class VI consists of stems that end with “Ayi”. The downloaded database which majorly consists of proper nouns does not contain any nouns ending with “Ayi” which generally occurs at the end of common nouns like “abbAyi” (boy), “ammAyi” (girl) etc. Class VII consists of only three stems “ceVyyi” (hand), goVyyi (pit), and “nuyyi (water well)” which end with “yyi”. All the three stems are thoroughly tested and one example is shown in section 5.1.

The evaluation results for oblique stem formation of the nouns are given in Table 21.

Table: 21 Evaluation Results of Oblique Stem Formation

Class	No. of nouns identified in each class
I	53
II	0
III	0
IV	0
V	35
VI	392

The evaluation results in Table 21 show that none of the nouns are categorized under Class II, Class III, and Class IV. Class II consists of non-human common nouns ending in “du”, “di”, “ru”, “ri”, “lu”, “li” which get replaced by “ti”. The downloaded dataset which majorly consists of proper nouns does not have any noun belonging to this class. Some

example nouns belonging to this class are listed in Table 9 of section 5.2.1. Class III consists of only six stems ending in “nnu”, and “llu” which are thoroughly tested (see Table 10 Section 5.2.1). The head of the complement in the example sentence (1) “iMtiki” shown in its root form “illu” in Figure 1 also belongs to Class III. Class IV consists of only seven stems ending in “yi”, and “rru” which are also tested thoroughly (see Table 11 Section 5.2.1).

Case marker generation primarily includes the oblique form generation and then joining the case marker to the form. As the results in table 21 show our engine generates the correct oblique forms. Our evaluation of case marker agglutination shows that for all the oblique forms our engine produces the correct surface noun forms.

The results show that all the 480 nouns were identified as belonging to different classes but in both the plural formation and oblique stem formation a few classes do not have any nouns in them. The list majorly consists of proper nouns and a very few common nouns because of which the nouns are not evenly distributed through all the classes.

9. CONCLUSION

The morphology engine for nouns and pronouns described in this paper along with a verb morphology engine are two separate modules in a surface realization engine for generating well- formed sentences in Telugu. We intend to look into the possibility of creating a generalized morphology engine for all word forms in Telugu language.

10. REFERENCES

- [1] Uma Maheswara Rao G and Amba Kulkarni P. 2006. Computer Applications in Indian Languages, Hyderabad: The Centre for Distance Education, University of Hyderabad.
- [2] Sasi Raja Sekhar Dokkara, Suresh Verma Penumathsa, and Somayajulu G. Sripada. 2015 A Simple Surface Realization Engine for Telugu
- [3] BH. Krishnamurti and J P L Gwynn, “A Grammar of Modern Telugu” Oxford University Press, 1985.
- [4] Albert Gatt and Ehud Reiter “SimpleNLG: A realization engine for practical applications”, Proceedings of ENLG 2009, pages 90-93, 2009.
- [5] Sri BadriNarayanan.R, Saravanan.S, Soman K.P “Data Driven Suffix List and Concatenation Algorithm for Telugu Morphological Generator” International Journal of Engineering Science and Technology (IJEST), 2009.
- [6] Reiter E. and Dale R. (2000), Building natural language generation systems, Cambridge University Press, New York.
- [7] Ganapathiraju M; Lori Levin. (2006): “TelMore: Morphological Generator for Telugu Nouns and Verbs”, in Proceedings of Second International Conference on Universal Digital Library, Alexandria, Egypt, pp. 17-19.
- [8] Hockett (1954): "Two models of grammatical description", in : Word, 10, pp. 210–234. [= Readings in Linguistics, vol. I, pp. 386–399].

- [9] Gregory T. Stump 2001: “Inflectional Morphology: A Theory of Paradigm Structure” Cambridge University Press.
- [10] Beesley, Kenneth R. and Lauri Karttunen. 2003. “Finite State Morphology”. Palo Alto, CA: CSLI Publications.
- [11] Lauri Karttunen and Kenneth R. Beesley 2005 “Twenty-Five Years of Finite State Morphology ”. Inquiries into Words, Constraints and Contexts.
- [12] Roark, Brain, and Richard Sproat. 2007. “Computational approaches to Morphology and Syntax” Oxford.
- [13] Lauri Karttunen 2003 “Computing with Realizational Morphology”.
- [14] Minnen, G., J. Carroll and D. Pearce (2000) Robust, applied morphological generation. In Proceedings of the 1st International Natural Language Generation Conference, Mitzpe Ramon, Israel. 201-208.
- [15] PJ Antony, KP Soman (2012): Computational morphology and natural language parsing for Indian languages: a literature survey. International Journal of Computer Science and Engineering Technology
- [16] Vishal Goyal, Gurpreet Singh Lehal (2011): “Hindi to Punjabi Machine Translation System” Proceedings of the ACL-HLT 2011 System Demonstrations, pages 1–6, Portland, Oregon, USA, 21 June 2011
- [17] Kristina Toutanova, Hisami Suzuki, Achim Ruopp (2008): Applying Morphology Generation Models to Machine Translation. Proceedings of ACL-08: HLT, pages 514–522, Columbus, Ohio, USA.
- [18] Bharti A, Chaitanya V, Sangal R “Natural Language Processing A Paninian Perspective” Prentice-Hall of India, New Delhi, 1995.