Interdisciplinary Journal of Linguistics
Volume [8] 2015, Pp.215-224

An analysis of Genitive in Telugu for Dependency Parser

Addanki Srinivas
K Rajyarama [*]

Abstract

In this paper, we focus on analyzing genitive case in Telugu, for the purpose of developing a syntactic parser. Genitive case, treated as falling under sixth case (shashti) based on the Panini's classification, denotes relation between two nouns, rather than between a verb and a noun. Therefore, it is also designated as aupa vibhakti 'sub-case.' (chinnaya Suri,1855) Although genitive is most often associated with possession, it is found to be performing a number of other functions as well. In this paper, we attempt to examine in detail the form and the various functions performed by genitive case in Telugu with a view to provide a rule-based system for machine learning purpose. The data is analyzed based on the Paninian Dependency Model. Paninian Dependency Model is considered the best model to analyse morphologically complex Indian languages.

Key words: Genitive case, Syntactic Parser, Rule-based system, Paninian Dependency Model

1.0 Introduction:

Genitive case is generally used to define, describe or classify the noun which it modifies. The relation between the genitive and the noun which it modifies can be varied and cannot be restricted to possession alone (Lander,Y,A. 2009). This is evident from the equivalent term for genitive *sesha shashti* in Indian grammatical tradition. The Sanskrit word *sesha* means *saMbaMdha* 'relation' and can be understood as encompassing all possible relations between the two nouns, while *Shashti* means 'six' and with reference to case system it refers to 'sixth case'. Case in Telugu is realized by suffixes (vibhakti pratyayas) being attached to the oblique form of nouns. Therefore case marking in Telugu involves two layers: In the first layer oblique stem formation takes place and it is to these oblique

[*] Centre for Applied Linguistics and Translational Studies, University of Hyderabad, Hyderabad, Telangana, India

stems/bases case suffixes get attached. The oblique stem functions as the sustaining form to bear case markers.[1] The following table illustrates this:
Table 1:

Sl.no	Base (Nominative)	Oblique stem	Case markers attached **(noun-obl-c.m.)**	Gloss
1	*kuuturu* 'daughter'	*kuuturi*	*kuuturi-**ni***	'daughter-acc'
2.	*guuDu* 'nest'	*guuTi*	*guuTi-**ki***	'to the nest'
3.	*ceyyi* 'hand'	*Ceeti*	*ceeti-**loo***	'in the hand'
4.	*vaaLLu* 'they'	*vaaLLa*	*vaaLLa-**too***	'with them'

However, oblique stem formation is optional and its realization is conditioned by the phonological shape of the base. It is interesting to note that the same noun in certain instances is realized in the oblique form while in some others it is not. For example, the noun *pustakaM* 'book' remains in the uninflected base (non-oblique) form for the attachment of locative and associative/instrumental cases: *pustakaM–loo, pustakaM-too* but realized as oblique for accusative and dative cases: *pustakaan-ni, pustakaani-ki.*
Genitive in Telugu is morphologically not overtly realized and therefore, for purpose of mapping be treated as null (ø) category. Consider the following examples:

1. T: ***kuuturi-*** ø *peLLi* 'daughter's wedding'
   Daughter (obl) wedding

2. T. ***guuTi-*** ø ciluka 'bird of the nest'
   Nest (obl.) bird

3. T. ***vaaLLa-*** ø pustakaM 'their book'
   They(obl.) book

Since genitive in Telugu does not have an overt realization, the oblique-ø form itself functions as the linking element between the two nouns in an adnominal. The formal similarity between oblique stem and genitive can be disambiguated

[1] List of symbols and abbreviations used in the paper: acc=accusative, c.m.=case marker, obl.=Oblique, gen=genitive, pl.=Plural,pst=Past, nh=Non-human, tv=Transitive verb, iv=Intransitive verb, nm=Non masculine, dat=Dative, NP=Noun phrase, 3p=3rd person, Mod aux=Mpdal auxiliary.

on the basis of distribution. Distributionally the modified nominal stem when inflected for case suffix or a postposition should be treated as oblique and when followed by another noun as genitive.

The present paper focuses on providing linguistic inputs for developing a Parser for Telugu. Parser involves analyzing language structure, in terms of syntactico-semantic relations operative at the sentential level. The Paninian Dependency Grammar that forms the base for analysis in this paper, treats a sentence as a series of modifier- modified relations. Each sentence has a primary modified category (generally a verb) and modifiers (the nouns). There are two levels of analysis in this model: A deep level of sematic relations and a surface syntactic level that has Relation markers (Vibhaktis). The deep level comprises of the direct participants of the action denoted by a verb (karaka) and other relations such as purpose, genitive, reason etc. The classificatory labels for the purpose of annotating the data are divided into two classes: k-Relations and non-k-Relations. All those nouns that have a direct relation with the verb fall under the former category while other categories though not desired by the verb directly, yet contribute crucially to the sentential semantics are placed under the second category. The following tags used for designating various kinds of relations illustrate this: k-Relations:k1=karta, k2=karma, k3=karaNa, k4=sampradana, k5=apadana. non-k-Relations: rt=tadardhya, r6=genitive, r6-k1,r6-k2 *karta or karma* of a conjunct verb(complex predicate) etc.

The paper is organized as follows: Section 2 gives details about the data and the methodology involved. Section 3 discusses the various semantic relations realized by the genitive case. Section 4 discusses the ambiguous structures and Section 5 concludes the paper.

2.0 Data

The data comprises of a set of 5k sentences drawn from web representing health, travel and tourism, literary and other domains of language use. We have analyzed and classified the data on the basis of the tag sets proposed by IIIT Hyderabad (cf. Akshara Bharati et.al, 2012). The syntactic and semantic cues provided by the IIIT Hyderabad guidelines (2012) have been taken as points of reference.

2.1 Procedure Involved: Annotating the data for dependency relations involves the following steps:

a. Corpus extraction from various sources
b. Corpus cleaning: The corpus obtained from different web sites is checked and cleaned of spelling and grammatical errors
c. The cleaned corpus is run through the modules of Tokenizer, Morph Analyzer and Chunker before being annotated.

2.2 **Distribution of Genitive:**

Genitive is found to be occurring mainly as a modifier, preceding the head noun in an adnominal construction in Telugu as in *ramuDi baaNaM* 'rama's arrow'. The head noun can either be a simple one or a derived deverbal nominal as in *ramuDi raaka 'rama's arrival'.* In the case of the head noun being a verbal derivative, the verb's argument structure remains intact, thereby retaining the modifying noun's role as the agent. This can be represented by adding the tag k1 to the existing tag r6 used to denote genitive.

3.0 **Controversy Surrounding the Use of old Telugu Genitive Marker *yokka :***

Although there is no overt morphological realization of the genitive in modern Telugu, old Telugu attests the use of a specific postposition *yokka* as genitive marker. Ex. *raamuDi baaNaM 'rama's arrow'* can also be expressed as *raamuDi yokka baaNaM '*arrow of Rama'. The use of *yokka* is treated as a remnant of old Telugu with no valid evidence as to its origin and etymology and as being found only in literary Telugu. (Krishnamurti, 2000 pp.63-67). One of the claims in this paper is that the use of *yokka* in certain contexts, despite the criticism, will help resolve the ambiguity raised by null realization of genitive. Evidences in support of this claim are provided in the following sections.

Genitive, in its function as possessive can also be marked by the case marker *-ki/-ku'* 'of', and postpositions *ceMdina* 'belongs to', *taaluukaa* 'belongs to'. The use of postposition *taaluukaa* 'belongs to' is restricted to a sub-set of possessive nouns, nouns which are alienable. For example, *vaaDi manushulu* 'his men' *vaaDi taalukaa manushulu '*men belonging to him' is possible but not **vaaDi taalukaa kaLLu* 'eyes belonging to him'. On the other hand, the use of the postposition *ceMdina* 'belonging to' would require the possessor noun to be marked for the dative case marker *-ki* making it a syntactic construct, denoting possession at clause level as in *vaaDiki ceMdina vastuvul* 'Things belonging to him.' The use of dative in Telugu to denote possession both alienable and inalienable has been attested earlier. (SubbaRao,K.V. and P.Bhaskararao, (2004), Uma Maheshwar Rao et.al 2012).

The alternant modes of expressing genitive denoting possession are exemplified below:

4. T: *naa tammuDu* 'My brother'
   - *a. naa-* ***ø-*** *tammuDu* = NP (obl.)+ø+NP
   - *b. naa* ***yokka*** *tammuDu* = NP (obl.)+ post pos.+NP
   - *c. naa-****ku*** *tammuDu* = NP (obl.)+dat.+NP

4.0 **Range of Semantic Relations Between a Genitive Noun and the Head Noun:**

As mentioned earlier (cf. section 1.0) the range of semantic relations between the two referents i.e. genitive noun and its head is vast including a number of

semantic domains like physiological, psychological, material etc. and should not be confined to possession. This section focuses on the possible relations between N1 and N2 in adnominal constructions.
The set of possible semantic relations between the constituent nouns in adnominals and the possibility of *yokka* insertion between the two nouns is illustrated in the following table:

Table 2: Semantic Relations

Semantic Relation	Noun-Noun compounds	Yokka Insertion
Possessive	*raamuDi pustakaM* rama-gen. book 'Rama's book'	✓
Possessive	*karNuDi radhaM* *KarNa-obl.- chariot* *'Karna's chariot'*	✓
Agent-creation	*devuDi srishTi* God-obl. creation 'God's creation'	✓
Agent-creation	*Nannaya bhaarataM* Nannaya-obl.-Bharatam 'Nannaya's Bharatam'	✓
Agency-vehicle	gurrapu baLLu Horse-obl.-carriage 'Horse carriage'	X
Source	*maTTi bommalu* *clay-obl.toys'* clay toys	X
Purpose	*paala ceMbu.* Milk-obl.-vessel 'Milk vessel'	✓
Quantitative: Age	*padeeLLa kurraaDu* ten years-obl.-boy' 'Ten year old boy'	X
Quantity: Measurement	ayidu miiTarla baTTa five metres-obl.-cloth 'five metres cloth'	X
Named entity-generic noun	*Dhillii nagaraM* Delhi-obl.-city 'the city of Delhi'	X
Part-whole	*iMTi kiTikiilu*	✓

	House –obl.-windows 'Windows of the house'	
Coordinate (numerals)	*muuDu vaMdala iravai aaru* three hundred-obl. twenty six 'Three hundred and twenty six'	X

From the above table it is possible to state that adnominals where the genitive denotes possession and part-whole (meronymy) relations permit insertion of yokka whereas those of purpose, quantity, quality, source do not. It can also be observed that the cases where *yokka* insertion is possible are also the ones in which the attributive noun is in the oblique form. Earlier studies on possession have tried to examine the relation between morpho-syntactic encoding of the possessor and possessum NPs and their semantics in terms of alienable vs. inalienable possession. (Nikolaeva & Andrew Spencer, 2010, Haspelmath, 2006). Telugu does not distinguish between alienable and inalienable possession, in terms of morphological encoding.

5.0 The problem:

The absence of an overt genitive marker and the optimality of oblique stem realization trigger the following problems:

a. When more than two nouns occur in a sequence it is difficult to predict which two of them enter into an adnominal relation.
b. When the two nouns involved are related to each other in more than one way.

Consider the following sentence where a number of nouns, which are uninflected for any case relation are placed in a sequence:

5. T: accaayamma^{N1} ulli pesaraTTu^{N2} vaasana^{N3} ravi^{N4} nooru^{N5} uureelaa ceesiMdi. Accayamma onion pesarattu smell Ravi mouth water- do-pst-3p-sg-nm.
'The smell of accayamma's onion pesarattu made ravi's mouth water'

In the above example five nouns appear in a sequence, all in their bare stem form without showing any overt case marking. Based on the animacy scale, it is possible to predict which two of them enter into relationship and the kind of relationship:

The animacy scale: human > animate > inanimate

As per the animacy scale humans rank above animals, followed by plants, natural forces, concrete objects, and abstract objects, in that order. In referring to humans, this scale contains a hierarchy of 'Persons', ranking first and second person pronouns above third person. Additionally, the hierarchy tends to place singular persons over plural.

In addition to the animacy scale, the possibility of *yokka* insertion will be of further help to resolve ambiguity:

6. T: accaayamma^{N1} ***yokka*** ulli pesaraTTu^{N2} ***yokka*** vaasana^{N3} ravi^{N4} ***yokka*** nooru^{N5} uureelaa ceesiMdi.

Ontological properties of the nouns that yield to *yokka* insertion confirm to the animacy scale:

N1 [+human] **yokka** N2 [+concrete, + artefact, +edible] N3 [-concrete, +olfactory], N4 [+human]

N5 [+concrete,+body part,+Part/whole].

The possessor is more frequently a human, animate entity while the possessum is lower on the scale of hierarchy. Therefore, it is possible to state that *yokka* insertion also complies with the animacy hierarchy. Morphologically, *yokka* insertion is related to oblique stem formation. It becomes obligatory when the N1 is in the bare uninflected (non-oblique) form and optional when N1 is realized as oblique stem. *Yokka* insertion helps resolve ambiguity and increases clarity.

Consider the following examples, where it is possible to have alternative readings between the two nouns:

7. T: prajala paalana
   People-pl-obl rule
   'Rule of the people'

Reading 1: *prajalu ceesee paalana* 'Rule by the people'
Reading 2: *prajala yokka paalana* 'Rule of the people'

Example (7) illustrates the issue raised in section 5.0 (b) where multiple relations between N1 and N2 are possible. The second reading is made possible with *yokka* insertion.

5.1 **Is adnominal Relation determined by verb as well?**

Genitive unlike other cases is not a dependent category on the verb. However the following examples appear to provide counter evidences to this fact:

8. T: *accaayamma^{N1} (yokka) reMDu^{NUM} pesaraTlu^{N2} unnaayi/baagunnaayi/maaDi pooyaayi*
   accayamma two pesarattus be-pst-nh-pl/good-3p-nh-pl/burn-pst-3p-nh-pl
   'accayamma's both the pesarattus are here/ good/got burnt'

9. T: **accaayamma^{N1} (yokka) reMDu^{NUM} pesaraTlu^{N2} veesiMdi.*
   accayamma two pesarattus- be-pst-nh-pl.
   'accayamma made two pesarattus'

The differential interpretation of N1in (8) and (9) is due to the verb in each one of these sentences. In (8) the verbs being intransitive require a single argument. Moreover ontologically the verbs *uMDu* 'Be' (existential verb), baagunDu 'be good' and maaDi poovu 'burnt' denote states thereby N1 functions as genitive and the relation between the two Ns is one of possessor and possessum.

While in (9) the verb *veyyi 'to make'*, a two-place predicate's argument-structure requirements are fulfilled by the two NPs (N^1 & N^2) in the sentence. The outcome is that N1 functions as the agent while N2 as the theme, confirming to the animacy hierarchy. This precludes the possibility of *yokka* insertion between the two nouns.

Similarly, the verb root when inflected for the modal auxiliaries –*aali* or –*valayu_vaccu* may lead to ambiguity:

10. T: *accaayamma* ^{N1} **(yokka) reMDu*^{NUM} *pesaraTlu* ^{N2} *tinaali/tinaalsi vacciMdi..*
    accayamma two pesarattus- eat-obl/ eat-happen-pst
    'accayamma had to /yet to eat two pesarattus'

Since the verb form is neutral for agreement, it is possible to have the following interpretations:

a. The verb is a transitive verb and N1 and N2 function to satisfy the argument structure requirements of the verb, thereby N^1 & N^2 has a sequential interpretation where N^1 is the *karta* (k1) and N^2N2 is the *karma* (k2).
b. The verb which by default shows agreement with karta (k1), is in the neutral form, enabling agreement either with N^1 or alternatively it is possible to make explicit an NP which is deemed to be missing via the pro-drop parameter. The N^1 -N^2 construct in this context has an adnominal interpretation.

10^1 T: *neenu/naaku accaayamma* ^{N1} *(yokka) reMDu*^{NUM} *pesaraTlu* ^{N2} *tinaalsi vacciMdi..*
I/to me accayamma two pesarattus- eat-obl-pst
'I had to eat accayamma's two pesarattus'

11. T: accaayamma^{N1} eekaika kumaaruDi^{N2} peLLiki veLLaali.
Accayamma only son-obl marriage-to go-obl
'accayamma has to go to her son's wedding'/ 'someone has to go to accayamma's son's marriage'

The ambiguity can be resolved by *yokka* insertion:

11^1 T: *accaayamma yokka eekaika kumaaruDi peLLiki veLLaali.*
Accayamma of only son-obl marriage-to go-obl
'x has to go for achayamma's son's marriage'

5.2 **Multiple Modifiers**:

When a number of modifiers precede the head noun, *yokka* insertion would bring about more clarity:

12. T: *utpattidaarula vinimayadaarula abhipraayaM*
    Producers-obl consumers-obl views
    'The views of producers and consumers (combined)'
13. T: *utpattidaarula yokka viniyoogadaarula yokka abhipraayaM*
    Prducers-obl- of consumers of views
    'The views of producers and consumers (individual)

The insertion of *yokka* would yield the reading in (13) as 'The individualviews of producers and consumers' as against (12) where it is 'The collective views of producers and consumers'.

Conclusion:

Given a sequence of any two nouns, confirming the status of the construction (adnominal or otherwise) is determined by the following factors:

1. Noun morphology
2. Verb's argument structure requirements
3. Modal auxiliaries
4. Verb semantics
5. Use of explicit genitive marker *yokka*
6. Animacy hierarchy
7. Prominence scale

N1-N2:

Table 3: **Analysis of N1 based on all parameters**

Obl.stem	cm	N2	Tv	Iv	Mod.aux	Pro-drop	Verb Ontology	Karta	Yokka Insert.	Gen.	Obl
✓	✓	X	X	X	X	X	X	X	X	X	✓
X	✓	X	X	X	X	X	X	X	X	X	✓
✓	X	✓	X	X	X	X	X	X	✓	✓	X
X	X	✓	X	X	X	X	X	X	✓	✓	X
X	X	✓	✓	X	X	X	X	✓	X	X	X
X	X	✓	✓	X	✓	✓	✓	✓	✓	✓	X
X	X	✓	✓	X	✓	X	✓	X	✓	✓	X
X	X	✓	X	✓	✓	✓	✓	X	✓	✓	X

Based on the above table the following rules help in identifying genitive in a N1-N2 construct:

Based on noun morphology:

1. $N1_{+/- obl.}$ + c.m. → Oblique stem

2. $N1_{+/- obl.}$- N2 → Genitive

Based on verb form and verb valence:

[N1-N2]

N1(+/- Obl.) –V_{tv}_mod.aux.{-aali/valayu_vaccu}→ k1(karta)

N1(+/- Obl.) – V_{iv}_ mod.aux.{-aali/valayu_vaccu}→ Genitive

Based on verb ontology:

[N1-N2] – $V_{\mathbf{iv}}$ [+existential, +Stative] -→ $N1_{\mathbf{Gen.}}$

[N1-N2] – V_{tv} [+any] → $N1_{karta.}$

Based on *yokka* insertion:

N1-yokka-N2 -→ $N1_{gen.}$

N1-N2 -→ $N1_{karta}$

Works of this nature which involve an in depth study of language structure would provide inputs for the development of Morpholgical Analyzers and also in POS tagging, Chunking and Transfer Grammar components.

References

Bharati A., Chaitanya V. and Sangal. R. 1995. *Natural Language Processing: A Paninian*
Perspective, Prentice-Hall of India, New Delhi, pp. 65-106.

Bharati A., Sharma D., Husain S., Bai L., Begam R. and Sangal R. 2012. AnnCorra: *TreeBanks for Indian Languages, Guidelines for Annotating Hindi TreeBank* IIIT, Hyderabad.

Chinnaya Suri ,P. (1855) *Balavyakaranam. Madras*: Vavilla & Sons.

Haspelmath, Martin. 2006. "*Explaining alienability contrasts in adnominal possession: economy vs. iconicity"*. Handout of invited talk at the Syntax of the World's Languages conference, Lancaster, 14-17 September 2006.

Krishnamurti, Bhadriraju. 2003. *The Dravidian Languages*. Cambridge et al.: Cambridge University Press (Cambridge Language Surveys)

Krishnamurti, Bhadriraju. 2000. *shashTiivibhakti 'yokka' vicaaraNa* in *bhashaa samaajaM saMskruti* Neelkamal Publications PVT.LTD., Hyderabad,pp.63-67.

SubbaRao,K.V. and P. Bhaskararao (ed.) (2004)*Non-nominative subjects in Telugu* . Volumes 1 &2. Amsterdam/Philadelphia: John Benjamins.

❐❐❐