Journal of Theoretical and Applied Information Technology
31st May 2018. Vol.96. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# A NEW FRAMEWORK : POST-POSITIONAL PARSER AND NALYZER FOR TELUGU PRONOUNS

[1]. P.PRASAD RAO [2]Dr S. PHANI KUMAR

[1] Research Scholar, GITAM University, Department Of CSE, INDIA
[2] Professor and Head, GITAM University, Department Of CSE, INDIA

Email : 1 prasadraopatwari@stmarysgroup.com 2 phanikumar.singamsetty@gitam.edu

**ABSTRACT**

**Objectives** : Postpositional Parser and Analyzer is a program which compiles and analyses pronouns of Telugu language into their roots and their constituent postpositions along with their grammatical properties, which is a full length translation surface of Telugu Pronouns . Grammatical properties at word level includes finding a tense, number, gender , person and other associated features. **Methods** :The method employed in this paper is based on Suffix-Order-Pattern technique to develop a computational model for morphological analysis of pronouns in Telugu. Suffix-Order-Pattern is a new technique in linguistic computation. The current work is a module of a POS Tagger and Word-Level translation surface for Telugu, a java application developed for analyzing Telugu pronouns. Test samples were taken from grammar text books for Telugu language and tested thoroughly with various inflections of pronouns with respect to tense, person, number and gender. **Findings:** The evaluation was performed on a data set taken from Wiktionary (Wiktionary is a multilingual, Web-based project to create a free content dictionary, available in over 150 languages). The results were very encouraging as the accuracy of the developed software is really absolute. The findings cannot be compared with any other results published because very little work was done previously in this area of research in Telugu language. **Application/Improvements:** The procedure used to build this is theoretically justified and practically executed for variety of inflections related to Telugu pronouns . The present proposal is a demonstration of XML based repository for all the root words of different categories of Telugu Pronouns. It is an optimal organization of linguistic database and its performance in computational environment is highly motivated..

**Keywords**: *Post-Positions, Root, Oblique, Suffix-Order, Case Markers, Inflections,*

## 1. INTRODUCTION

### 1.1 Pronoun Analyzer

Telugu is a Dravidian language spoken by people from the south Indian states of Andhra Pradesh and Telangana. It is a morphologically rich free word order language with nearly 90 million first language speakers[15]. In this paper we describe a Pronoun Morphological Analyzer (MA) engine which automatically analyses the different forms of pronouns in Telugu. Morphological Analyzer (MA) and Morphological Generator (MG) are two very important parts of Natural Language Processing (NLP) applications like machine translation systems and surface realization engines[15]. A Morphological Analyzer analyses a given word and processes it into its root along with its grammatical information whereas a Morphological Generator given a root along with its grammatical information generates the corresponding word[15]. The current work is a separate module of a Word-level Translation Surface engine for Telugu, a framework implemented for analyzing different pronoun forms of Telugu language.

### 1.2.Machine Translation

Machine Translation mentions the use of computers to convert some or the entire task of translation between human languages[16]. Automatic translation between languages regarded as a complex task due to the variety of grammar rules and different syntactical phenomena in literature. Various implementations of machine translation have been proposed in computational

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

linguistics. The rule-based approach is the first strategy in Machine Translation that was developed. A Rule-Based Machine Translation system consists of collection of rules, called grammar rules, a bilingual or multilingual lexicon to process the rules[16].

Translation approach involves the process of analyzing morphological information, identifying the constituents and record the words in the source language according to the word order pattern of the target language and then replace the words in the source language by the target language words using a lexical dictionary of that particular language. Simplifying and segmenting an input language text becomes mandatory in order to improve the machine translation quality. Development of a machine translation (MT) system requires very close collaboration among linguists, professional translators and computer engineers.

The grammar should be optimized with a view to obtaining a single correct output[13 ].Grammatical descriptions of human languages are the results of efforts in modeling of the design features and the internal organization of the structures and the mechanisms of language. Therefore, Linguistics is about language modeling, designing and studying their theoretical and practical implications. However the activity of grammatical descriptions itself is modeled by the specific functional aspects of language and Linguistic Computing[17] .

### 1.3.Telugu to English translation:

Telugu is an Indian language spoken by over 100 million people in the states of Telangana and Andhra Pradesh. The language is a highly structured, disciplined and rich in terms of expression, style and construction[15].The richness of Telugu language lies in the extremely large number of words representing different moods, expressions, contexts[15].

Till now, most of the research done for English language computational model. Due to the morphological richness of Indian languages, researchers faced a great difficulty to write complex linguistic rules in terms of computational methods. Process models for Indian languages was mainly based on rule based approaches[17].Rule-based method requires expert linguistic knowledge and hand written rule. This means that the set of rules must be properly written and checked by human experts. NO such automation is available for Telugu language at word level.

### 1.4 About Morphology

Morphology has been well studied both by theoretical and computational linguists. From a theoretical perspective ,structure of explained by the following three models.

Item and Arrangement model which is morpheme based Morphological approach in which word forms are analyzed as arrangement of morphemes. In this model a morpheme treated as the minimal meaningful unit of a language and word are treated as concatenation of morphemes.

Item and process model which is lexeme based morphology in which a word form is assumed to be a result of applying rules that alter a stem to produce a new one. In this model inflectional rules ,derivational rules and compounding rules are applied to a stem to obtain required word form.

Word and paradigm model which is a word based morphological approach which states generalizations that hold between different forms of inflectional paradigms.

## 2.RELATED WORK

Work in the area of Machine Translation in India has been going on for several decades. During the early 90s, advanced research in the field of Artificial Intelligence and computational Linguistics made a promising development of translation technology. This helped in the development of usable Machine Translation Systems in certain well-defined domains..Many organizations like IIT(Kanpur), CDAC(Mumbai), CDAC(Pune), IIIT(Hyderabad) etc. are engaged in development of MT systems under projects sponsored by Department of Electronics(DoE), state governments etc. since 1990[13].

**ANGLABHARTI (2001**) -R M K Sinha, Jain R, Jain A developed a machine aided translation system designed for translating English to Indian languages.

**Shakti (2003)** - Bharati, R Moona, P Reddy, B Sankar, D M Sharma and R Sangal have developed a system which translates English to any

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Indian languages with simple system architecture[12].

**MAT(2002)**- Murthy K developed a machine assisted translation system for translating English texts into Kannada, which used morphological analyzer and generator for Kannada[12].

**Sampark**-In 2009 a machine translation system among Indian Languages was proposed by the Consortium of Institutions (IIT Hyderabad, University of Hyderabad, C-DAC Noida, Anna University, KBC Chennai, IIT Kharagpur, IISc Bangalore, IIIT Allahabad, Tamil University, and Jadavpur University). Currently released systems are {Punjabi, Urdu, Tamil, Marathi} to Hindi and Telugu-Hindi and Telugu to Tamil machine translation system[14] .

**Google Translator** : In 2007, Franz-Josef Och applied the statistical machine translation approach for Google Translate from English to other Languages and vice-versa. Hindi and Urdu are the only Indian Languages present among the 57 Languages for which Google Translate provides translation. Accuracy of the system is good enough to understand the sentence after translation [14].

English-Telugu machine translation system is developed jointly at CALTS with IIIT, Hyderabad, Telugu University, Hyderabad and Osmania University, Hyderabad. This system uses English-Telugu lexicon consisting of 4200 words[14 ].

## 3.PRONOUN ANALYZER

### 3.1 Mechanism.

This paper deals with a mechanism which provides a translation surface of Telugu pronouns from Telugu to English .This approach is based on rule based methodologies defined in the language. Telugu language contains well defined grammatical framework at word level, phrase level and sentence level with finite set of rules. This mechanism extracts word level grammatical description for different categories of pronouns in Telugu language. Grammatical description includes variety of analysis finding a root ,post position, number, gender and concerned inflection. This implementation accepts any Telugu pronoun as input and gives word level grammatical segmentation ,which becomes a translation surface and useful in automated translation from Telugu to any other language.

This mechanism includes variety of analysis finding a root, post position ,number, gender and concerned inflection for a given word which is a pronoun compound . Due to many practical issues involved in maintaining a database with all combinations of compounds, it is better to maintain only standard or root words. Compounds of the source language are split to obtain the original words using reverse engineering in accordance to the conjunction (‘sandhi’) rules. This will make the morphological analysis easier. This plays an important role in translation also whereas understanding is also a kind of translation. Splitting of compounds into root words is an important phase in natural language processing for the applications like machine translation . Building a computational model to analyze natural language is the goal of natural language processing .

Parsing is a challenging task for computer science engineering in case of natural languages because these languages are having different structures and different characteristics ,which makes analysis of a word or sentence too complex. In translation very commonly source is to be translated into target pattern. The word presented in any language to be analyzed in each context such as syntax, semantics , grammar and sense. The idea is meaningful units to be parsed systematically to obtain meaningful grammatical description. Way of working out these combinations, has probably been the most important theme in analysis.

### 3.2 About Telugu Pronouns

Pronoun is a function word that is used in place of a noun or noun phrase .

Ex:- **నేను**,**మేము**, **అతడు**, **ఆమె** etc.

**Classification of Telugu pronouns**

According to common classification pronouns are divided into eight types. This classification facilitates to build efficient computational framework for Telugu pronouns.

Journal of Theoretical and Applied Information Technology
31st May 2018. Vol.96. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 1- Types of Pronouns*

| Pronoun Type | Example |
|---|---|
| Personal-Demonstrative pronouns | **నేను/మేము** |
| Reflexive pronouns | తాను/తాము |
| Interrogative pronouns | ఎవడు/ఎవరు |
| Indefinite pronouns | ఎవడైన/ఎవరైన |
| Demonstrative adjective pronouns | ఆ/ఈ/ఏ |
| Possessive-adjective pronouns | నా/నీ/ మా |
| Pronouns-referring to number | ఎందరు/ ఎన్ని |
| Distributive pronouns | ఒక్కొక్క |

All above pronouns generate different inflections with variety of grammatical properties defined in the language,

Nouns and pronouns have both direct and oblique stems. Direct Stems are used in the nominative case. The oblique stem marks the genitive and is also base for the remaining cases which are marked by adding suffixes to it.

*Table-2 Direct-Oblique stems*

| Direct(nominative) | Oblique(other form) |
|---|---|
| నేను | నా |
| నీవు | నీ |
| వాడు | వాని |
| వారు | వారి |
| అతడు | అతని |
| ఆమె | ఆమె |
| తాను | తన |

Pronouns in Telugu like in many languages often exhibit a considerable amount of irregularities when inflected for number (Singular, plural),gender(masculine, nonmasculine ),person (1st, 2nd, and 3rd) , case (nominative, accusative, genitive, dative, vocative, instrumental, locative) and different post positions. This sort of behavior is generating large number of words which are belonging to pronoun category. Storing various pronoun inflections directly in any data store derives complexity while analyzing words independently and also it is a tedious task in computation..This proposed technique optimizes the data store with a linguistic based classification which is mapped to computational phenomena.

*Table-3 Case Marker-Suffixes*

| Case Marker | Suffixes |
|---|---|
| Nominative | డు, ము, వు, లు |
| Accusative/Objective | నిన్, నున్, లన్, కూర్చి, గురించి |
| Instrumental | చేతన్, చేన్, తోడన్, తోన్ |
| Dative | కొఱకున్, కై |
| Ablative | వలనన్, కంటెన్, పట్టి |
| **Genitive** | **కిన్, కున్, యొక్క, లోన్, లోపలన్** |
| Locative | అందున్, నన్ |

### 3.3.Computational Framework

**This** framework focuses on how linguistic based phenomena of Telugu pronouns will be mapped with computational model which efficiently handles the processing of different inflections of pronouns and provides an optimized translation surface for Telugu-English translation.

Machine translation is such a complex task because of various reasons and it happens at sentence level or at paragraph level. Telugu is a source language and replaced with English language.

Journal of Theoretical and Applied Information Technology
31st May 2018. Vol.96. No 10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

## 4. IMPLEMENTATION

### 4.1 Classification of Inflections:

#### 4.1.1 Inflections with Number, Person, Gender

a)Inflected for Number(singular-plural)

Ex:-Singular-**నేను**(Nominative)
plural – **మేము**(Nominative)

b)Inflected for person(First, Second, Third)

Ex:- First person - నేను
Second Person- నీవు
Third person - అతడు

c)Inflected for Gender(Masculine non-Masculine)

Masculine - అతడు(Nominative)
Non Masculine - ఆమె(Nominative)

Pronouns when inflected for number, person, gender are generating limited number of inflections those can be stored directly in data store with nominative-oblique structure.

#### 4.1.2.Inflections with Post Positions:

Post positions are the words take suffix pattern while inflecting Telugu pronouns. These are the words inflecting pronouns in various ways. In this context linguistic based phenomena to be structured as computational framework with an optimized and meaningful classification of different inflectional patterns These postpositional inflections can be categorized .into two types.

a) Inflections with case markers.
b) Inflections with adverbs.

#### 4.1.2.1 1Inflections with case markers.

Variety of inflections occur with case markers in case of pronouns are categorized based on number, person and gender for different classification. These inflections are generating large number of pronoun words ,storing all these words in database is a tedious task and processing the words stored in a database becomes a complex task in computation. The proposed technique stores nominative and genitive case for the entire classification of pronouns. Nominative represents the direct form and genitive represents the oblique from.

Ex:- 1.వారితో= వారి + తో

Oblique form of Pronoun : వారి
Case marker : తో {Instrumental case}
2. నాకంటె= నా+ కంటె
Oblique form of Pronoun : నా
Case marker : కంటె{Ablative case}

#### 4.1.2.2 Inflections with Adverbs.

Telugu adverbs are inflecting pronouns and generating large number of pronoun inflections. All these adverbs will be stored in a data storage repository as post positions.

Ex:- 1. అతనైన= అతను+ అఇన
Pronoun : అతను
Post position : అఇన is an adverb.

2. తమరేమో= తమరు+ఏమో
Pronoun : తమరు
Post position : ఏమో is an adverb.

### 4.2 Data store Representation:-

#### 4.2.1 Data store Optimization:

Optimization is a technique which leads to efficient computation while processing the data resides with any format of data storage. This proposal focuses on data store optimization with a linguistic based classification. Pronouns in Telugu are visible in different forms with different inflections.

This representation stores the required data in an XML file which is tree structured often referred to as XML Tree and plays an important role to describe any XML document easily. The tree structure contains root (parent) elements, child elements and so on. By using tree structure, traversal of all succeeding branches and sub-branches starting from the root can be done very easily. The parsing starts at the root, then moves down the first branch to an element, take the first branch from there, and so on to the leaf nodes.

Data store optimization enforces us to classify data storage into three parts.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**4.2.1.1 Root- Oblique,xml** - A root is a morho-lexical term for the morpheme of word that is most central in terms of morphology and semantics. The root is the morpheme that the other morphemes attach to.The root is the part that is not a suffix,prefix or infix. It stores all the root words and its oblique forms . This representation starts with an element pronoun type as parent moves to person and then it includes number and gender also. At most internal node consists of either three or four element references to extract the word level description and it also gives the root word and its oblique form.

*Table-4 Root-Oblique.Xml*

```
//third person demonstrative pronouns

<pronoun type="demonstrative">
<person type="third">
<dt vacanamu="esingular">
<dte gender="masculine">
<dtem root1="vadu">
<oblique>vani</oblique>
</dtem>
<dtem root2="vIdu">

<oblique>vIni</oblique>
</dtem>
```

At most internal node :<dtem root1=”vadu”>

Ex:- Pronoun type : <d***>Demonstrative
Person :<*t**> Third
Number:<**e*>Singular
Gender :<***m>Masculine

**4.2.2.2 Case-Markers.xml**

This file stores all the suffix reflections of seven Telugu case markers as root words in xml file. These root words have to be stored with a feasible pattern as these words are directly reflecting the splitting process when an input is pronoun word with case marker inflection.

This representation starts with an element case marker and includes attribute name as type which is a parent node , then it moves to a child reference which denotes the case marker type and includes suffix-order attribute to denote the number of characters in a case marker suffix . Internal node consists of an element reference which extracts the case marker type and number of characters in a suffix ,all the suffixes of different categories will be stored as root words with different xml notations .

Ex:-At most internal node :<I2 root1=”du”>

Case marker type :<I*>Instrumental
Suffix-order:<I2>
Instrumental number. of characters=2

*Table-5 Casemarkers.Xml*

```
<casemarker type=”Instrumental”>
<I order=2>
<I2 root1=”ce”> </I2>
<I2 root2=”wo”></I2>
</I>
<I order=4>
<I4 root1=”woda”></I4>
<I4 root2=”cewa”></I4>
</I>
</casemarker>
<casemarker type=”Nominative”>
<N order=2>
<N2 root1=”du”> </N2>
<N2 root2=”mu”></N2>
<N2 root3=”vu”></N2>
<N2 root2=”lu”></N2>
</N>
</casemarker>
```

**4.2.1.3 adverbs.xml:** stores all the post positions which are adverbs and inflecting pronouns in various ways. This representation facilitates splitting process as these post positions are inflecting pronouns at rear end. When an input is a pronoun word this post position to be identified directly with any feasible classification. Suffix - order pattern classifies different post positions occur as inflections ,suffix-order denotes the number of characters in a post position.

This representation starts with postposition as a parent and includes suffix-order attribute which denotes the number of characters in a post position. Then it moves to a child node and At most internal node consists of either three or four element references to extract the word level description and it also gives th e root word and its oblique form.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 6-Postpositions.xml*

```
<postposition suffix-order="">

    <ppa root1="aaina">
    <ppa root1="aaiweV"">
    <ppa root1="pE">
    <ppa root1="aeVmo">
     </ppa>
</postposition>
```

**4.2.4 Procedure to extract analysis of pronouns:**

Inflections are called 'postpositions' which play an important role in Telugu grammar. In Telugu, inflections occur at the rear part of a word which leads in altering the original form of the root word. If the word is inflected then it is not possible to carryout splitting straightaway. All inflections must be separated and splitting is applied to obtain root words. Conjunction is possible not only with root words but also with inflected words. For example, 'nAwo' is the compound which is inflected ('wO') at rear end. It is separated first and split rule is applied to obtain 'nA' + 'wo'. .

**4.2.4.1 Finding Postpositions:**

Postposition is a word added at rear end of pronoun. If any inflected word is given as input it is to be divided into two parts, one is oblique form of a pronoun and another one is postposition.

Ex:-nAwo= nA + wo (wx notation)

నాతో = నా + తో (Telugu word)

**Add**ing e suffixes to the oblique form derive large number of inflections in various formats. These formats can be categorized on the basis of suffix- order pattern .

Suffix-Order=No. of characters in suffix;

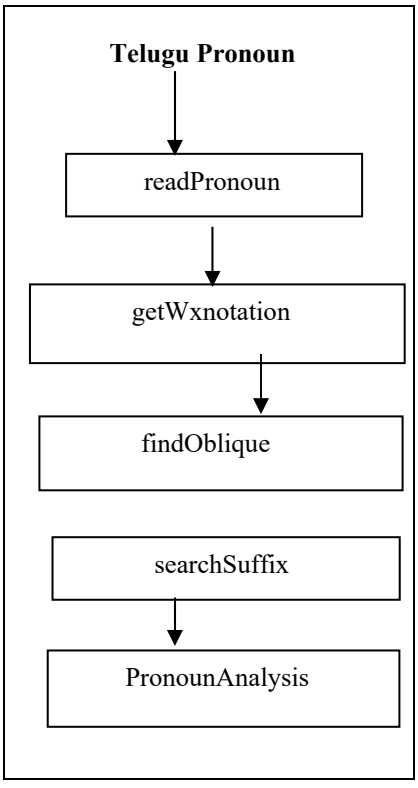
This pattern is a reference for splitting a pronoun word or pronoun phrase.

**4.2.4.2 Architecture and Framework Input Specification**

Pronoun analyzer in the current work is a part of POS Tagger and Word level Translation surface engine which compiles and analyses pronons of Telugu language into their roots and their constituent postpositions along with their grammatical properties and other associated features. The input for pronoun analyzer is a Telugu pronoun compound which contains any kind of valid inflection as per the rules specified in language domain and presented with WX notation.

Example : 1.awanu (He)
2.aAmeV (She)
1.awaniki (to him)
2.aAmewoV (with her)

The notation used in examples from 1 to 4 is to specify Telugu words in English is called as WX notation16. It is a very popular transliteration scheme for representing Indian languages in the ASCII character set. This scheme is widely used in Natural Language Processing in India.



ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Figure 1: Parser And Analyzer -Architecture Diagram*

This framework consists of five functional implementations and it uses one interface.

1.String readPronoun(String)
2. String getWxnotation(String).
3..String ndOblique_Postposition(String)
4.boolean searchSuffix(String:suffix)
5.void xtract_PronounAnalysis(Oblique)

**readPronoun(String)** : This implementation reads a string which is a Telugu pronoun from the user and returns the same to the method.

getWxnotation(string) :It converts Telugu pronoun into WX-Notation format.

WX-Notation is one well known format which replaces Telugu Alphabets with English Alphabets.

Ex;- 1. అ - a
2. ఆ - aA

**findOblique_Postposition(Teluguprono un) :** This functional implementation accepts the Telugu pronoun as a parameter which is of wx_notation format from application layer and proceeds with further execution. Objective of this function is, it has to find Oblique form and Postposition of given pronoun..It requires a computational phenomena which can split the given pronoun and segments it as two parts ,one is oblique and other one is post position.

Post positions are inflecting pronouns at rear end, the proposed computational phenomena implements a splitting process which starts at rear end of a string with a technique called Suffix-order pattern. This pattern classifies different post positions occur as inflections at rear end of pronoun ,suffix-order denotes the the number of characters in a post position. Value of Suffix- order pattern in this technique is greater than or equals to two according to the postpositional data storage formats. The process of splitting starts at the Suffix-order equals to two it means the last two characters of a string will be captured and this suffix existence will be verified with data storage repository.

Suffix existence with data storage will be verified with another functional implementation **searchSuufix(suffix)** ,if this implementation returns true the process reaches end state else the suffix order value will be incremented by one and the process continues as mentioned for suffix-order equals to two, it continues like this until searchSuffix(suffix) returns true state. Once this method returns true value implies splitting of Postposition is successful and method returns the oblique form of a pronoun. In one special case if SearchSuffix(suffix) returns false after all iterations it raises an exception as no suffix is existed only pronoun without any inflection is existed.

**searchSuffix(String:suffix):This** implementation checks Suffix existence in data storage repository which in xml format. The Document Object Model is an official recommendation of the World Wide Web Consortium, It defines an interface that enables programs to access and update the contents of XML documents. The DOM API defines the mechanism of searching an element with XML file which is used here to search the suffix in XML files. If search is successful method returns true else it returns false.

**Void extract_PronounAnalysis(Oblique):It** accepts Oblique form of pronoun and this method implementation constructs a phenomena of searching an oblique form in oblique-root, xml file. It uses DOM API to access and search the oblique form in xml file, once the search is successful the process returns entire morphological information of a pronoun in terms of case, gender, person and type.

## 5. RESULTS AND DISCUSSION

The developed software suite was rigorously experimented with variety of pronoun inflections. This application is evaluated for two main factors. The first one is proper splitting of a postposition from given pronoun and second one is extracting grammatical description includes finding the root word, person, gender, number and oblique .

Finite data set is constructed to store all types of telugu pronouns with proper classification, which is evaluated with all possible postpositional inflections. Postpositional corpus is also constructed with two data sets one is case markers data set and second one is adverbs data set and both these data sets are also finite.

There were 126 root words with their 126 respective oblique forms categorized into eight groups were taken and tested successfully with different combinations of postpositions.

Test cases for pronoun inflections are generated for each pronoun type with all possible combinations ,each combination contains postpositional inflection.

Example:-
No of Demonstrative Pronouns in corpus=42
No. of postpositions in corpus =40
No. of inflection Test cases=42*40

Four thousand seven hundred and twenty inflections are tested successfully for all the combinations of pronouns, out of this one specific case is observed and handled as an exception for which no postposition is found. There were 118 specific cases existed without postposition all these cases are tested successfully with absolute accuracy .

**Test Cases and Applications:**

Different test cases are generated to check the performance of the designed framework for various categories of telugu pronouns and associated inflections. Input is always a pronoun of Telugu language and this analyzer provides a surface for machine translation.

Machine translation includes all possible classifications of source word to be translated into destination language. This engine will construct a possible classification for telugu pronouns which are at source side.

**Examples:**

1.Telugu pronoun : manakMteV :

**Classification of given pronoun:**

mana → Oblique form
mana → root word
{ gender, person, number}→{not applicable, second, plural}

kMte → postposition

Application : Translation to English

Postposition will be considered first for given case and translation pattern is : PPOR{postposition, oblique, root}
After dictionary look up this will be translated as:

**manakMteV →Than us**
{ Telugu } {English }

Translation is successful**.**

2.aAmeVwoV :

**Classification of given pronoun:**

aAmeV → oblique form
aAmeV→ root word
{ gender, person, number}→{non masculine, third, singular}

woV → postposition

Application : Translation to English

Postposition will be considered first for given case and translation pattern is : PPOR{postposition, oblique, root}

After dictionary look up this will be translated as:

**aAmeVwoV→with her**
{ Telugu } {English }

Translation is successful**.**

3. .awanu :

**Classification of given pronoun:**

awani → Oblique form
awanu → root word
{ gender,person,number}→{ masculine ,third, singular}

Postposition- no

Application : Translation to English

Postposition will be considered first for given case and translation pattern is : NPPOR{no-postposition, oblique, root}

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

After dictionary look up this will be translated as:

**Awanu → He**
{ Telugu } {English }

Translation is successful**.**

*Table-7 Categorization of pronoun inflection Test cases*

| Group | Description | Number of test cases |
|---|---|---|
| 1 | Demonstrative | 42*40 |
| 2 | Reflexive | 24*40 |
| 3 | Interrogative | 16*40 |
| 4 | Indefinite | Formed with Inflection |
| 5 | Demonstrative adjective/interrogative adjective | 4*40 |
| 6 | Possessive adjective | Formed with Inflection |
| 7 | Referring to number | 12*40 |
| 8 | Distributive | 20*4 |

*Table-8 Sample test cases- Output Structure*

| I Input: Telugu pronoun | Post-position | Grammatical description |
|---|---|---|
| నాతో | తో | Oblique: నా Root: నేను |
| తమరికై | కై | Oblique: తమరి Root: తమరు |
| ఎవరైన | ఆఇన | Oblique: ఎవరి Root: ఎవరు |

The results were very encouraging as the accuracy of the developed software is really absolute.

Various pronoun compounds of Telugu language Can be translated to any destination, not only English language. All the grammatical features were listed with pre defined properties at language level. This surface can be marked as pronoun translation surface of Telugu language ,which is a source language for any destination.

Listing all grammatical properties were successful with this engine at different levels of word compound. Always post position will be recognized at first level, then oblique form of the pronoun is identified with defined grammatical properties. All the grammatical properties are pre defined in XML file called pronouncorpus.

## 6. CONCLUSION

In this paper a new approach of postpositional parsing is introduced and grammatical description for pronouns is extracted for all the parameters required for machine translation . It provides a surface for machine translation from Telugu to any other language . variety of inflections are captured and analyzed using this technique successfully.

This analyzer is providing a translation surface form source to destination with extracted grammatical description. It is generating different translation patterns like PPOP,NPPOP for translating any given telugu pronoun into destination language. Well defined classification at word level is successful with this engine.

Well structured and optimized data store representation is done to store roots and oblique forms of all Telugu pronouns and finite set of postpositions are stored in repository to identify pronoun inflections. Various categorical inflection

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

occurrences are found within this finite set only, no inflection is found beyond this set.

Suffix-Order Pattern is a new technique which is used for word splitting gave good results because all possible patterns are stored and traced successfully to derive postpositions. In one case if pronoun does not contain postposition suffix-order becomes zero and no postposition is derived ,only grammatical description will be extracted for such case . No test case is raising ambiguity while splitting pronoun as no need to write any rule to avoid ambiguity in this issue The proposed technique is well suited for finite data sets ,even this technique can be used for Telugu verbs and adjectives in future.

## REFERENCES

[1] A Telugu morphological analyzer -Uma Maheshwar Rao G., Amba Kulkarni P. and Christopher Mala Center for Applied Linguistics and Translation Studies University of Hyderabad ,Hyderabad, India

[2] Combining pos taggers for improved accuracy to create telugu annotated texts for information retrieval- Rama Sree, R.J.1, Kusuma Kumari P.2

[3] TelMore: Morphological Generator for Telugu Nouns and Verbs-Madhavi Ganapathiraju and Lori Levin Language Technologies Institute, Carnegie Mellon University, Pittsburgh, USA

[4] A Progressive Grammar of the Telugu Language with Copious Examples and Exercises-Arden,Albert Henry ,Society for promoting Christian knowledge. Madras

[5] A Simple Rule-Based Part of Speech Tagger-Eric Brill ,Department of Computer Science University of Pennsylvania Philadelphia, Pennsylvania .

[6] Annotating Corpora Guidelines For POS And Chunk Annotation For Indian Languages-Akshar Bharati, Dipti Misra Sharma, Lakshmi Bai , Rajeev Sangal Language Technologies Research Centre ,IIIT, Hyderabad.

[7] IIIT-Hyderabad: A Part-of-Speech Tagset for Indian Languages. http://shiva.iiit.ac.in/SPSAL2007/iiit_tagset_guidelines.pdf

[8] Anusaaraka: A better approach to Machine Translation { A case study for English-Hindi/Telugu} Akshar Bharati , Amba P Kulkarni , Dipti Misra Sharma Language Technologies Research Center, International Institute of Information Technology, Hyderabad .

[9] Part-of-Speech Tagging and Chunking with Maximum Entropy Model-Sandipan Dandapat, Department of Computer Science and Engineering, Indian Institute of Technology , Kharagpur, India 721302,

[10] Parts Of Speech Tagging for Indian Languages: A Literature Survey Antony P J Research Scholar Computational Engineering and Networking (CEN), Research Centre, Amrita Vishwa Vidyapeetham University, Coimbatore, India . Dr. Soman K P , Professor and Head Computational Engineering and Networking (CEN), Research Centre, Amrita Vishwa Vidyapeetham University, Coimbatore, India.

[11] Key Issues in Vowel Based Splitting of Telugu Bigrams: T. Kameswara Rao Assoc. Professor and Head, CSE Dept Brahma's Inst. of Engg. and Tech Rajupalem, Nellore, AP, India. Dr. T. V.Prasad Former Dean of Computing Sciences, Visvodaya Technical Academy, Kavali, AP, India.

[12] Survey of machine translation systems in India: G V Garje and G K Kharate Department of Computer Engineering and Information Technology PVG's College of Engineering and Technology, Pune, India Principal, Matoshri College of Engineering and Research Centre, Nashik ,India

[13] Machine Translation Systems for Indian Languages Latha R. Nair School of Engineering Cochin University of Science and Technology. David Peter S School of Engineering Cochin University of Science and Technology

[14] Survey of Indian Machine Translation Systems Sitender, 2Seema Bawa1,2Dept. of CSE, Thapar University, Patiala, Punjab, India

[15] Telugu to English Translation using Direct Machine Translation Approach T. Venkateswara Prasad1, G. Mayil Muthukumaran2 1Dean of Computing Sciences, Visvodaya Technical Academy, Kavali, AP, India 2Technical Director, National Informatics Centre, Govt. of India, New Delhi, India(1tvprasad2002@yahoo.com,2muthu@nic.in)

[16] Translation of Telugu –Marathi and vice-versa using rule based machine translation Dr. Siddhartha Ghosh1, Sujata Thamke2 and Kalyani U.R.S3 Head of the Department of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Computer Science & Engineering, KMIT, Narayanaguda, Hyderabad siddhartha@kmit.in 2R&D Staff of Computer Science & Engineering, KMIT, Narayanaguda, Hyderabad sujata.thamke@gmail.com 3R&D Staff of Information Technology, KMIT, Narayanaguda, Hyderabad upadhyayula.kalyani@gmail.com

[17] English-Telugu Rule Based Machine Translation system A Thesis submitted for the degree of Master of Science (by research) in the School of Engineering By R.Sri Badrinarayan