

## Rule Based Syntax Analyser for Telugu Poems

Smt.M.Humera Khanam

Associative Professor  
Department of  
Computer Science & engineering  
Sri Venkateswara University college of Engineering  
,Thirupati  
*e-mail:humera\_svec@yahoo.co.in*

Mr.I.Reddy Sekhar Reddy

M.Tech II YEAR  
Department of  
Computer Science & engineering  
Sri Venkateswara University college of Engineering  
,Thirupati  
*e-mail:irsr560@gmail.com*

**Abstract**--Natural language processing concerns about interpretation of language between human and machine such a way that machine can understand human language and communicate with a human so that human can understand. Challenging topics in this field is processing the Indian languages . The objective is to reduce human effort and time taken to perform tasks , improve the throughput and quality of a service. Chandhassu is a basic information to know whether a given poem is syntactically correct or not .Some poems are meaning full even though they doesn't follow Chandhassu but those are not good when listening and reading .Nowadays so many peoples are writing poetry but those poems are not popular as much as ancient poets due to the lack of knowledge about Chandhassu and vyakaranamu. Proposed system is help full to check whether a given poem is syntactically correct or not. This system can be developed based on predefined rules for each type of Chandhassu .In Sanskrit Chandhassu types are called vruthas, some vruthas are adapted to write telugu poetry, Vruthas based on number of letters. In telugu Chandhassu types are called jaathulu and upajaathulu . these are dependent on number of ganalu. And also it checks whether poem follows yathi and prasa properties . Poetry can be useful to represent large amount of information in a small and structured format such a way that it can be easy to remember. Main applications of Chandassu is to write a poems, songs, prabhandas (collection of poems), dwipada (poems that follow a typical pattern), Sataka (collection of 100 poems) etc. By using poetry we can able to effectively represent sthaya bhavulu (symptoms) . It is helpful to represent science of Chandhassu ,vyakarana ,alnkara ,sahithyamu . Poetry is also used to represent dictionaries , medicine ,math's , testing of vasthu ,pearls ,authentication purpose in data transition.

**Keywords:** Chandhassu, ganalu, tokenization, yathi, prasa, labels, vruthalu, jathulu, upajathulu.

\*\*\*\*\*

### 1. Introduction:

Natural Language processing (NLP) is a field of computer science and linguistics concerned with the interactions between computers and human (natural) languages. Language processing applications will use knowledge about language .NLP algorithms are based on machine learning, especially statistical machine learning. It requires knowledge on different fields like linguistics, computer science, statistics (particularly Bayesian statistics), linear algebra and optimization theory.

Research in natural language processing has been going on for several decades dating back to the late 1940s. Machine translation (MT) was the first computer-based application related to natural language. In the case of telugu language processing it is very difficult to process do to structure of the language.in this paper we will focus on syntax analysis. The study of structural relationships between words .This level focuses on analyzing the words in a sentence. This requires both a grammar and a parser.Knowledge need to order and group words together comes under a syntax. The most frequent applications

utilizing NLP include the following Information Retrieval , Information Extraction (IE) , Question-Answering Summarization , Machine Translation and Dialogue Systems.

Telugu is a one off the Dravidian language. It is the official language of Andhra Pradesh, Telangana states of India. It is also spoken in some other states like Karnataka, Tamil Nadu, Orissa, Maharashtra and Chattisgad. Telugu is the second most-spoken language in India (85 million native speakers according to the 2011 census).Globally there are 110 million peoples are speaking. There are so many forms of Telugu literature (**Telugu Sahityam**), widely divided as prose and poetry. The sub-forms in prose include - novels, short stories. The poetry is classified into poems, songs, prabhandas (collection of poems), dwipada (poems that follow a typical pattern), Sataka (collection of 100 poems), Avadhana Sahityamu , Asukavitha . In writing a poems poet can able to select appropriate Chandhassu based on the context this is called Chndhashilpam . this is very use full in the case of avadhanam and asukavitha .

Science of Chandhassu tells about characteristics of poems. It will come from a word called as chadhissamvarani. Characteristics of poems is called Chandhassu do to arranging an information in limited number of letters in particular order and that will very delight to lesion. This Chandhassu will be used from very past years ,it will be used in vedhas .all the Manthras present in rugvedham are followed Chandhassu .it is a one among the six divisions of vedhas those are 1: siksha 2:vyakaranamu 3:chandhu 4: nirukthamu 5:jyothishamu 6:kalpamu . In telugu literechur poetry can be not only used in kavyas, prabandhas , sathakas but also used in science of Chandhassu ,vyakarana ,alnkara ,sahithyamu .poetry is also used to represent dictionaries ,physics ,medicine ,math's ,testing of vasthu ,pearls.

In Sanskrit some corers of vruthas are their among those mostly below hundred vruthas are regularly used in writing a slokas ,among this hundred only five to ten vruthas are used to write a telugu poetry all this vruthas are dependent on “ya maa tha raa ja bha na sa la gam “ ganalu.

This paper comes under syntax analysis , it tell about whether a given poem syntactically correct or not,based on predefined rules that are mentioned in science of Chandhassu .

## 2. System Architecture:

It will take input as a telugu poem , check whether a given poem follow any type of Chandhassu rules or not .If it follow then it will produce which type of Chandhassu it will followed ,letter used for prasa if it present,letters used in yathimaitri places. If it will not follow any Chandhassu then it will produce output as syntax error in a given poem. Architecture will shown in figure 1 .

It will take input as a poem then send it to deciding of lagu or guru stage.In this stage it will count number of lines ,each line divided into words , each word divided into letters then each letter labelled with either lagu or guru .That labelled data will send to labelled with ganalu. Here labelled data will divided into ganalu ,each ganam will labelled with appropriate letter from “ya ma tha ra jab ha na sa la gam”. This data will send to Chandhassu mapper it will identify type of Chandhassu and return to user and also send it to recognizer of yathi and prasa letters.itwill take poem and poem type as a input and finally it produce prasa letter ,yathi mithri placed between the letters.

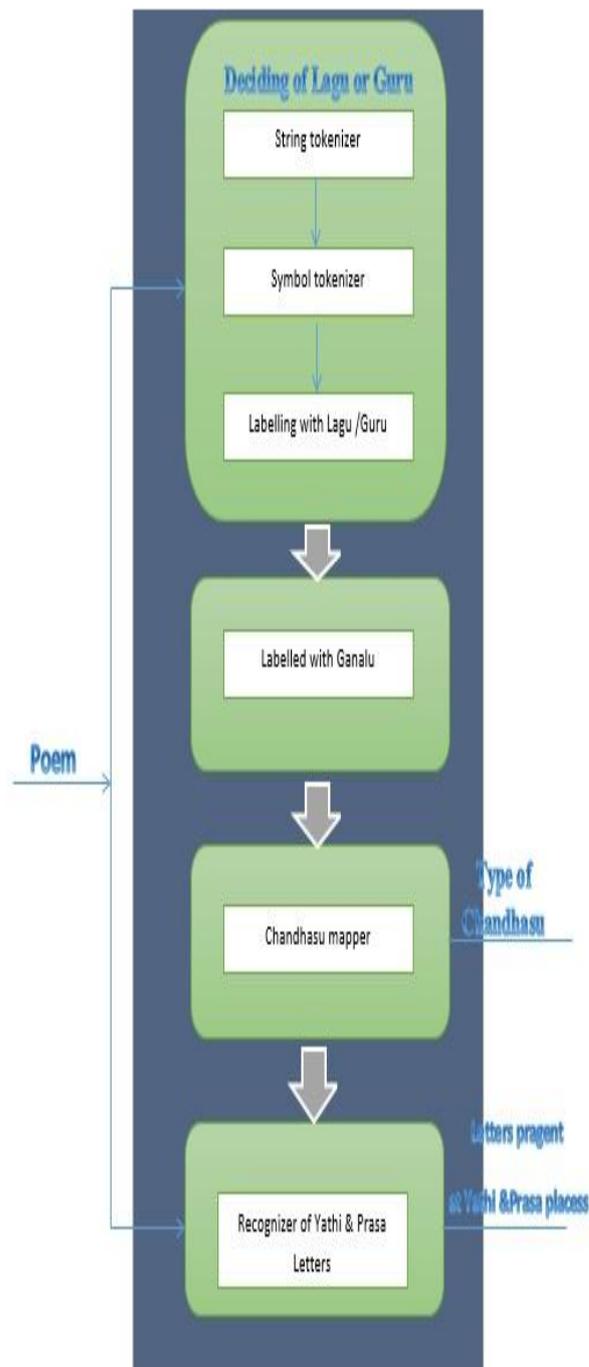


Fig1:Architecture of a Rule based syntax analyzer

## 3. Design of syntax analyser :

Syntax analyser contains four important components those are 1.Deciding of lagu or guru , 2.labelled with ganalu ,3. Chandhassu recogniser ,4. Recognizer of yathi and prasa letters.

### 3.1. Deciding of lagu or guru:

This component will take input as a poem finally it will give out as labelled data i.e each letter in a poem is

labelled with either laghu or guru. For this module input space is {all the telugu letters} output space is{ U,I }. Internal process of this component is initially poem can be divided into words based on word separator. Hear word separator is ‘-’, ‘,’,’?’,’!’.Those words can be divided into letters based on letter separator. Hear letter separator is space. For deciding each letter either lagu or guru based on some conditions those are all dhirgaksharalu ,letters present with nidusuna will make it as a guru. Before letter for dhithvaksharalu ,samukthaksharalu will make it as a guru. Letters having polluhallu or visrga then that letter will labelled with guru. After this all the remaining letters will make it as a laghu.

Example:

Input poem:

సారధియంట-వేదముల-సారము-శౌరి-తదంఘ్రి-భక్తి-చె

న్నారెడు-క్రీ డి-తా-రధికుడానట-చిందము-విల్లు-దేరునన్

వారువముల్-మొదల్-దివిజవర్గ-మొసంగినవంట-యస్త్రవి

స్తారగురుల్-శివాదులట-సంగరమం-దెవడాగ-జాలెడిన్

Inshially poem will send to sting tokenizer hear poem can divided into words

సారధియంట	వేదముల	సారము	శౌరి-తదంఘ్రి
భక్తి	చె		
న్నారెడు	క్రీ డి	తా	రధికుడానట
చిందము	విల్లు		
దేరునన్			
వారువముల్	మొదల్	దివిజవర్గ	మొసంగినవంట
యస్త్రవి			
స్తారగురుల్	శివాదులట	సంగరమం	
దెవడాగ	జాలెడిన్		

Then it will send to symbol tokenizer hear each word divided into letters .

సా	ర	ధి	యం	ట
వే	ద	ము	ల	
సా	ర	ము		
శౌ	రి			
త	దం	ఘ్రి		
భ	క్తి			
చె	న్నా	రె	డు	

క్రీ	డి	తా			
ర	ధి	కు	డా	న	ట
చిం	ద	ము			
వి	ల్లు				
దే	రు	నన్			
వా	రు	వ	ముల్		
మొ	దల్				
ది	వి	జ	వ	ర్గ	
మొ	సం	గి	న	వం	ట
య	స్త్ర	వి			
స్తా	ర	గు	రుల్		
శి	వా	దు	ల	ట	
సం	గ	ర	మం		
దె	వ	డా	గ		
జా	లె	డిన్			

It will send to labelling with lagu or guru hear each letter will labelled with lagu or guru base on the algorithm proposed in ‘algorithm 1’.

Algorithm lagu-guru-mapping(p)

```

{
//p-poem
//da-dirga,nidusuna,visargaksharalu
//su-samuktha,dithwakshralu
//ph-pollu hallu aksharalu
//l-number of lines in poem
//m-number of words in each line
//n-number of letters in each word
for i:=0 to l
{
for j:=0 to m
{
for k:=0 to n
{
for a:=0 to x
{
if (p[i][j][k]=da[a]) then return ‘U’;
else return ‘I’;
}
}
for a1:=0 to x1
{
if(p[i][j][k]=su[a1])

```

```

    {
        if(j=0 & k=0 & i!=0) then
return 'U' as a previous line last letter;
        else if k=0 then
            continue ;
        else
return 'U' as a previous letter;
    }
    else
        continue ;
    }
    for a2:=0 to x2
    {
if(p[i][j][k] ends with ph[a2]) then
        return 'U';
        else
            continue;
    }}
}
}

```

**Algorithm 1:**labelling with either lagu or guru  
Eg: U I I U I U I I I U I I

సారథియంటపేదములసారము

U I I U I U I U

శౌరితదంఘ్రిభక్తిచె

U I I U I U I I I U I I

న్నారెడుక్రీ డితారధికుడొనట

U I I U I U I U

చిందమువిల్లుదేరునన్

U I I U I U I I I U I

వారువముల్మోదల్లివిజవర్గ

I U I I U I U I U

మొసంగినవంటయస్త్రవి

U I I U I U I I I U I I

స్తారగురుల్పివాదులటసంగర

U I I U I U I U

మందెవడాగజాలెడిన్

### 3.2 Labelled with ganalu:

Hear labels of each line will divided into ganalau. In vruthas each ganamu have three labels final ganamu have remaining labels in each line. But in the case of jathulu and upajathulu will depends on character sticks of poem and ganalu that will used i.e either surya or indhraganalu this may contain either two or three or four labels for each ganamu. Hear we will divide and labelled with ganalu for vruthas only. By using algorithms proposed in algorithm 2 will divide and labelled with Ganalu. This labeling of all vruthas ganalu can be done based on “యమాతారాజభానసలగం” ganalu .యగణం-“IUU”, మగణం-“UUU”, తగణం-“UUI”, రగణం-“UIU”, జగణం-“IUI”, భగణం-“UII”, నగణం-“III”, సగణం-“IIU”, వగణం-“IU”.

Algorithm division-labelling-of-ganalu(pglg);

```

{
//pglg-program labelled ganalu
//sg-standardganalu{"I","U","U","U","I","U",
"I","I","I","U"}
//lsg-labels of standard ganalu{"య","మ" , "త" , "ర" ,"జ"
,"భ" ,"న" ,"స","ల","గం" }
if (nofl=4)
{
If same number of letters in each line then
rm:=mod(nofletters,3);
x:=nofletters-rm;
for i:0 to noflines
{
for j:0 to x
{
for kl:0 to 8
{
If(pglg[i][j]=sg[kl]& pglg[i][j+1]=sg[kl+1] &
pglg[i][j+2]=sg[kl+2])
{
Then return lsg[kl];
break;
}
else
{
continue;
}
}
}
j:=j+3;
}
if(rm=1)
{
if(pglg[i][x]="U")

```

```

then return “గ”;
}
else if(rm=2)
{
if(pglg[i][x]="" && pglg[i][x+1]=""U”)
then return “వ”;
} }
}
}

```

**Algorithm 2:**division and labelling of ganalu.

Eg:

```

భ ర న భ
U I I | U I U | I I I | U I I
భ ర వ
| U I I | U I U | I U
భ ర న భ
U I I | U I U | I I I | U I I
భ ర వ
| U I I | U I U | I U
భ ర న భ
U I I | U I U | I I I | U I I
భ ర వ
| U I I | U I U | I U

```

### 3.3 Chandhasu mapper:

It will take input as labels of ganalu. Based on labels it will map to appropriate Chandhasu by using algorithm proposed in algorithm 3.

భ ర న భ ర వ-ఉత్పలమాల

న జ భ జ జ జ ర-చంపకమాల

మ స జ స త గ-శార్దూలం

స భ ర న మ య వ-మత్తేభం

న జ జ జ జ జ వ-కవీరాజవీరాజితము

Above all are vruthas those ganalu are in same order in four lines of a poem .This characteristics will referred from [1].

Algorithm chandhasumapper(lganalu)

```

{
//lganalu-labels generated by program for a given poem
//utpm-labels of ganalu for utpalamala
//mtbm-labels of ganalu for mathebham
//chmp-labels of ganalu for champakamala
//srdl-labels of ganalu for shardhulam
//kvrv-labels of ganalu for kavirajavirajitham
if(nofletters=20)
{
if(lganalu=utpm) then return ఉత్పలమాల
else if(lganalu=mtbm) then return మత్తేభం
}
else if(nofletters=21)
{
If(lganalu=chpm) then return చంపకమాల
}
else if(nofletters=19)
{
If(lganalu=srdl) then return శార్దూలం
}
else if(nofletters=23)
{
If(lganalu=kvrv)
then return కవీరాజవీరాజితము
}
}
}
Algorithm 3:Chandhasu mapper

```

Eg: ఉత్పలమాల

### 3.4: Recognizer of Yathi & Prasa Letters:

Hear it will check second letter of each line have same hallu or not .if it present means given poem fallows prasa condition and return prasa letter used in a poem.

Eg:ఓ

For checking yathimithri letters it will take input as poem and type of Chandhasu. Based on type of Chandhasu yathimithri letters will change. uthpalamala will fallow 1<sup>st</sup> letter of each line with 10<sup>th</sup> letter ,champakamala fallows 1<sup>st</sup> letter of each line with 11<sup>th</sup> letter ,mathebbham fallows 1<sup>st</sup> letter of each line with 14<sup>th</sup> letter ,shardhulam fallows 1<sup>st</sup> letter of each line with 19<sup>th</sup> letter ,kavirajavirajitham fallows 1<sup>st</sup> letter of each line with 8<sup>th</sup> ,14<sup>th</sup> letters. All this rules referred from [1] and [2].

Eg: achulu halulu

(అ-అ) ( స-స)

(అ-అ) ( న-న)

(అ-అ) ( వ-వ)

(అ-అ) ( స-స)

In each letter only one acchu will present but more than one hallu is present. Among this we can make any hallu will yathimithri . in a given example 4<sup>th</sup> line 1<sup>st</sup> letter contains స and త butyathimithri will maked to స only.

### 4. Evaluation :

For evaluating our system we have taken poems from reference [3],[4], [5], [6], [7], [8] ,[9] and [10] and given to a system it given a accurate result for all the poems of five types.we will given nearly 500 poems among all five types.Here we are considering one measurement for evaluating our system:*success rate*.

evaluation measure defined as follows:

success rate in % =(Number of poems that will map correct chandhasu by the system/ Number of poem given by the user)\*100 %.

### 5. Conclusion:

This syntax analyser will fallows a rule based approach, so if a give poem fallows rules of Chandhasu mentioned in reference [1] it will give exact results otherwise it will give a given poem have syntax error and also display given poem is nearer to witch type of Chandhasu where you need to modify.

Further extension for this system is to prediction of appropriate word in writing a poem that will suitable to specified Chandhasu and also it will give same meaning as mention word.

### References:

- [1] Thelugu Bhashalo Chandhorethulu By Ravuri Dhorasamisharma
- [2] Sulaxana Saramu By Limgamagumta Thimmakavi

- [3] Thelugu Vyakharanamu By Raghavendhra Publications
- [4] Paamdavodyogam By Thirupathi Venkata Kavulu
- [5] Srimadhandhra Mahabharatham By Kavithrayamu
- [6] Srimadbhagavatham By Pothana
- [7] Parijathapaharanamu By Nandhi Thimana
- [8] Vasucharithra By Ramarajabhushanudu
- [9] Vijayavilasamu By Chemakura Venkataakavi
- [10] Amukthamalyadha By Srikrishnadevarayalu
- [11] Bharthruhari Subhasithalu By Yenugu Lakshmanakavi
- [12] Radhikasaamthwanamu By Muddhupalani
- [13] Introduction To Machinelearning By Ethem Alpaydn
- [14] Speech And Language Processing By Daniel Jurafsky And James H.Martin
- [15] Maya Report On Telugu Litrechure .