International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

Rule Based Machine translation of Complex Sentences from English to Telugu

K Deepthi Krishna Yadav & L Keerthi Lingam
[1]M.Tech Student, Dept. of C.S.E., Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam.
[2] Assistant Professor, Dept. of C.S.E., Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam.
kdkrishnayadav.15.cstmtech@anits.edu.in & keerthi.cse@anits.edu.in

ABSTRACT: *Machine translation is one of the most developing and research oriented topics of Computational Linguistics. Researchers aim to develop a Machine Translation system that produces high quality translation outputs with high accuracy, efficient parsing and covering maximum language pairs. Translation of complex sentences is one of the major challenges in Natural Language Processing. Our proposed approach deals with Dictionary-based Bi-Lingual (Uni-directional) machine translation. We focus on identifying the 'clauses' and 'Subordinate conjunctions' which play an important role in simplifying and translating the complex sentences by preserving the meaning and the structure of the Source text (ST). This paper discusses the results obtained by implementing the algorithm in a machine translation system on a sample test suites.*

KEYWORDS: MT: Machine Translation, SL: Source Language, CL: Clause Identification, CR: Clausal Re-ordering, SC: Subordinate- Conjunction positioning, TL: Target Language

I. INTRODUCTION

Language is a system of human communication, either written or spoken, by a group of words in a structural and a conventional way. It is thought to have its origin from when early hominins gradually started modifying their primate communication systems, acquiring the ability to form the theory of learning other minds and shared intentions. This development made many linguistics see the structures of language to serve for specific communications and social functions. Language, of course, is referred to as Knowledge, and in our today's industrial developing world, knowledge plays as a key factor for competitiveness which creates growth and development with prosperity we tend to take for granted. In India, major languages being spoken belong to Indo-Aryan languages which constitute of 75% and the Dravidian languages spoken by 20%of Indians. Other languages belong to Austroasiatic, Sino-Tibetan, Tai-Kadai and a few other minor language families. **Translation** is derived from the Latin word *translatio*, meaning "*a carrying across*" or "*a bringing across*": in this case, of a text from one language to another. According to Western theories, If translation be an art, and it is not an easy one. If the translation is to be true, the translator should know, understand and master both the languages, as well as the science that he/she is to translate. Machine translation is a computerized process of converting a Source Language (SL) from one human language to the other known as Target Language (TL) by preserving not only the meaning, the style, form and effect of the Source Language (SL). Research in machine translation has been going from 50 years and is still developing. Machine translation works well for Simple sentences while faces difficulty in translating compound and complex sentences which are connected by means of connectives comprising of primary and secondary conjunctions. In order to overcome the difficulties faced in translation of complex sentences, Sentence Simplification is performed by identifying the clauses present in the complex sentence there by re-ordering them and translating them by using several approaches such as Rule-based, Data-driven, Phrase-based approach etc.

1.1 Complex Sentences

Complex sentences are formed by connecting two unlike clauses i.e.., Dependent and the Independent clauses. Here the connectives used for joining the unlike clauses are Secondary conjunctions or Subordinating Conjunctions (SC). These connectives represent the relation between the clauses.

Independent clause: The Independent clause can stand alone as a sentence giving a complete thought or sense. It is also referred to as main clause.

Dependent Clause: The Dependent clause cannot stand alone though it has a subject and verb. It depends on the main clause (Independent clause) to express a complete idea. If it appears at the beginning of a sentence, it is usually followed by comma. In case if it follows the main clause, punctuation is not mentioned.

II. RELATED WORK

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

In [1] authors presented adaptive rule based machine translation from English to Telugu. This approach is based on rule-based methodologies. If-then methods to select the best rules for target language in translation, Probability based appropriate word selection for a given sentence and rough sets to classify a given sentence are the approaches used in this technique. Set of production rules of English and Telugu, Training set and Dictionary for both the languages are developed for this purpose. In [2] authors presented a "Relative clause based Text simplification for Improved English to Hindi Translation. In this approach they are focusing on the identification of the Relative clause for text simplification. In [3] authors presented an architecture for a text simplification system based on previously developed general coverage tools (giving them a new utility) and on hand written rules specific for syntactic simplification. These rules are based on morphological features, focused on specific phenomena like appositions, finite relative clauses. The simplification could be used for both humans and machines. This is the first proposal for "Automatic Text simplification and opens a research line for the Basque language" in NLP. In [4] authors developed an approach in which they applied a Rule-Based for text simplification and improving the quality of Gujarathi-Hindi translation system. According to their approach, the Source text re-arranged themselves according to the structure of the target text. They have evaluated the quality in terms of BLEU, NIST, mWER and found that there approach ranked first in terms of simplicity, fluency and adequacy.

III. Translation Perspective Of Complex Sentence

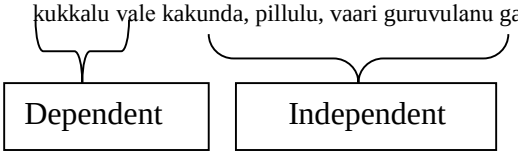
We know that English is a Fixed-word order language which has the flexibility in the position of the clauses while Telugu is a Free-word order language which is rigid in its clauses. In Telugu, a subordinate clause always appears at the beginning of the complex sentence. The Independent clause(IC) always follows the dependent clause (DC) of the sentence. In English, a subordinate clause (SC) appears at any position of the complex sentence. The different Structures of complex sentence in Telugu are as follows:

Type 1: Structures of complex sentence in Telugu:

Ex: SL: Cats, unlike dogs, do not respect their masters.

TL: కుక్కల వలె కాకుండా, పిల్లులు, వారి గురువులను గౌరవించవు.

kukkalu vale kakunda, pillulu, vaari guruvulanu gauravinchavu.



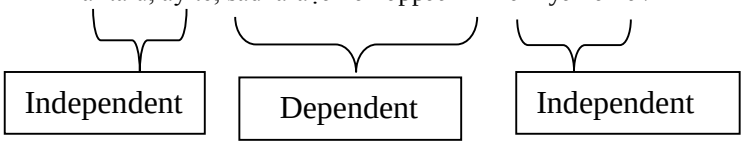
(1) "Unlike" is the Subordinate Conjunction(SC) in the above sentence.

Type 2: Structure of complex sentence in Telugu:

Ex: SL: The results, however general, are important.

TL: ఫలితాలు, అయితే, సాధారణమైనప్పటికి ముఖ్యమైనవి.

Phalitālu, ayitē, sādhāraṇamainappatiki mukhyamainavi



(2) "However" is the Subordinate Conjunction (SC) in the above sentence.

IV. Subordinate Conjunctions (SC) As Connectives

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

A Conjunction is a word that joins or connects together words, phrases, clauses, or sentences. There are mainly two types of conjunctions, i.e.., Primary class of *Coordinating conjunctions* and a secondary class called *Subordinating or Subordinate conjunctions.* Words called Conjunctive adverbs sometimes act like conjunctions, but other times act like old adverbs.

The **subordinate conjunctions** are used to connect unlike clauses i.e.., Dependent and the Independent clauses by means known as Connectives. In English, Connectives always appear at the **beginning** of the Dependent clause (DC). They represent how the Dependent clause (DC) is related to the Independent clause (IDC). The Subordinate Conjunctions are represented by syntactic tags SC, these can be prepositions (IN), adverbs (RB), Wh–pronouns (WP), Wh- adverbs (WRB).

Table 1. SC's as connectives and their translation in English and Telugu.

Condition	Comparison	Cause	Parallel	Place
		English		
Even if, incase, unless, whether	Though, Although, whereas, while, than, who	As, Because, Since, So that, than, how	After, before, still, till, until, when, whenever, while	Where, Wherever
కానీ, చో, ఎడల	ఐనను, ఐనప్పటికీ	నట్లు, కైచే, చెత	వరుకు, తరువాత, పిమ్మట,	ఎక్కడ, ఎక్కడైన
Kani, Cho, Edala	Ainanu, Ainappatiki	Natlu, Kai, Che, Chetha	Varaku, Taruvatha, Pimmata, Appudu	Ekada, Ekadaina

V. PROCEDURE FOR TRANSLATING COMPLEX SENTENCES

Machine translation systems includes the basic steps for lexical, syntactic and semantic analysis like POS tagging, chunking, parsing, re-ordering and generation. Apart from these some key activities were involved during the translation of complex sentences. These components include Clause Identification (CL), Clausal Re-ordering (CR) and Subordinate Conjunction (SC) positioning.

Clausal Identification: It is the process of separating the independent and the dependent clauses in a complex sentence. The lexical, syntactic and semantic analysis is performed for each clause separately. Here, the sentence is simplified by means of a delimiter (',') wherever present and identifying the position of the Subordinate Conjunction in the sentence.

Ex: S: The movie, though very long, was very enjoyable.

S1: The movie was very enjoyable.

S2: though very long

Clausal Re-ordering: Generally, the semantic order of the words in a sentence varies from language to language. In Telugu, the Subordinate Conjunction (SC) always appears at the beginning of a complex sentence. The dependent clause precedes the independent clause of the complex sentence. Whereas in English, the Subordinate Conjunction appears at any position, i.e.., it can be present at the beginning followed by an independent clause or it can follow an independent clause.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

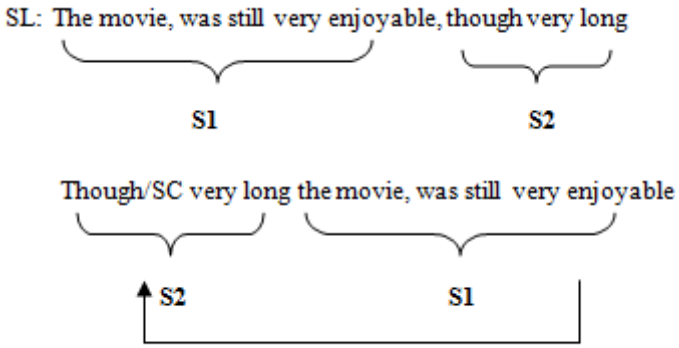
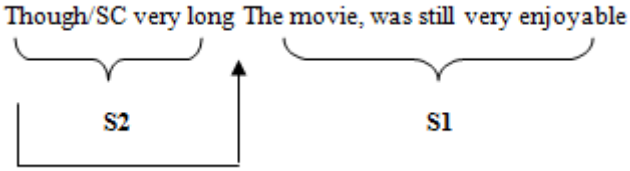


Fig: Clausal re-ordering

Subordinate Conjunction (SC) positioning: In Telugu, the Subordinate conjunction (SC) appears at the end of the dependent clause occupying the place of an object. Hence, the SC should be moved to the end position of the Dependent Clause (DC).



TL: Very long *though*, the movie was very enjoyable.
చాలా కాలం అయినప్పటికి ఆ చలన చిత్రం చాలా

VI. ALGORITHM FOR TRANSLATING COMPLEX SENTENCES

The algorithm is developed to handle translation of complex sentences separated by one or more commas containing a single Subordinate Conjunction in the sentence from Source Language (SL) to Target Language (TL).

Consider two empty arrays S1 and S2. S1= Independent clause S2 = Dependent clause

```
Step 1: Get the Source Input sentence.
Step 2: Split the Input sentence into an array of strings, 'S' using the delimiter space
        (' ')
Step 3: for i=0 to S. length
        {

         if (S[i] ! = Subordinate Conjunction (SC))
                {
                   S2[ ] = S[i];
                }

         else
              Select all the routes which have active nodes
                 {
                          S1 [ ] = S [ i ];
                           i++;

                            while ( S[i] doesn't contain (,))
                           {
                           S1[ ] = S [ i ];
                            i++;
```

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

```
                    }

                      S1 [  ] = S[ i ];

                }
        }
```

Step 4: Applying Subordinate Conjunction (SC) shift

```
        For i = 0 to N
        {
            If (S[ i ] = = SC)
                {
                    array _push(S1, S1 [ i ] );
                     unset (S1 [ i ] );
                }
        }
```

Step 5: Implode (S1);
Step 6: Implode (S1);
Step 7: array _merge (S1, S2).
Step 8: Translate (S1), Translate (S2)

VII. SIMULATION RESULTS

The algorithm was implemented to translate complex sentences in English to Telugu machine translation system. . To evaluate the system performance in translating complex sentences formed using connectives a test suite of around 30 sentences was developed. The Machine Translation system was tested on the test suites and the outputs generated by the system were generated based on adequacy and fluency measures. The exact translations of the sentences were prepared with the help of linguistic experts of SL and TL.

Source Language	**Subordinate Conjunction movement (SC Movement**	**Target Language (Rule-based Translation performed)**
After our trip from the beach, School started back, and I was excited to see my friends.	our trip from the beach *after* School started back and I was excited to see my Friends	సముద్రం నుండి మన యాత్ర తరువాత పాఠశాల తిరిగి మొదలయ్యింది మరియు నాకు సంతోషంగా ఉంది నా స్నేహితులను చూడుటకు.
The movie, though very long, was still very enjoyable.	very long *though* the movie was still very enjoyable	చాలా సమయం అయినప్పటికీ ఆ చలన చిత్రము ఇప్పటికి చాలా ఆనందంగా ఉంది.
Cats, Unlike Dogs, do not respect their masters.	Dogs *Unlike* Cats do not respect their masters.	కుక్కల వలె కాకుండా, పిల్లులు, వారి గురువులను గౌరవించవు.
The Results, however general, are important.	General *however* the results are important.	ఫలితాలు, అయితే, సాధారణమైనప్పటికి ముఖ్యమైనవి.
The Book, even though lengthy, concept is very interesting.	Lengthy *even though* the book concept is very interesting.	పుస్తకం, సుదీర్ఘమైనప్పటికీ, భావన చాలా ఆసక్తికరంగా ఉంటుంది.

SCREENSHOTS OF OUTPUTS:

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

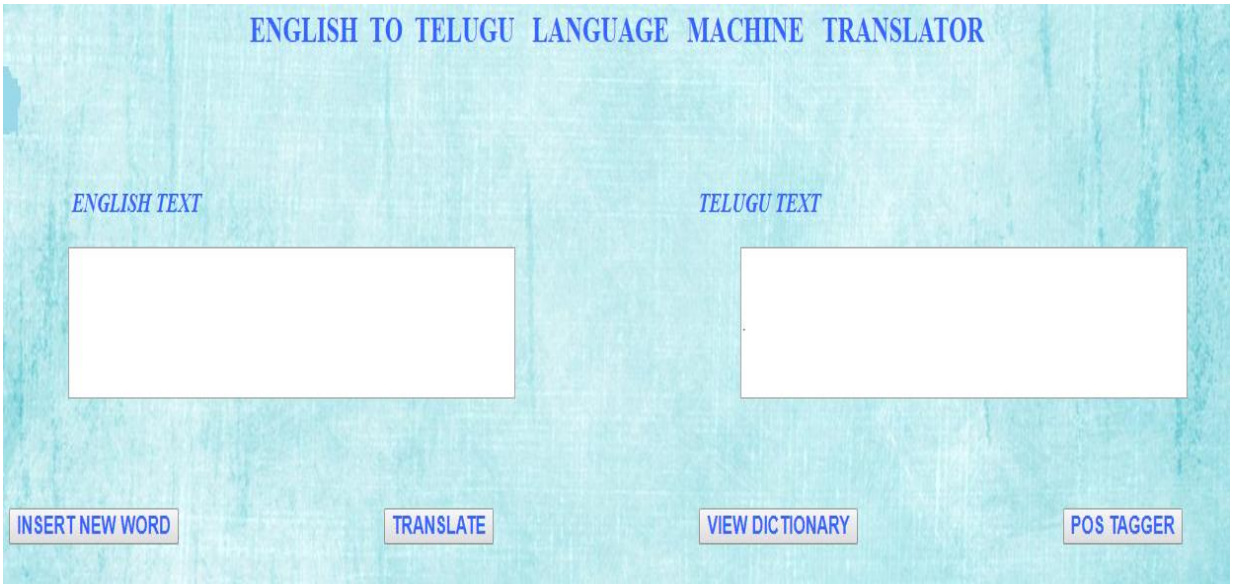


Fig1. Translation Workspace

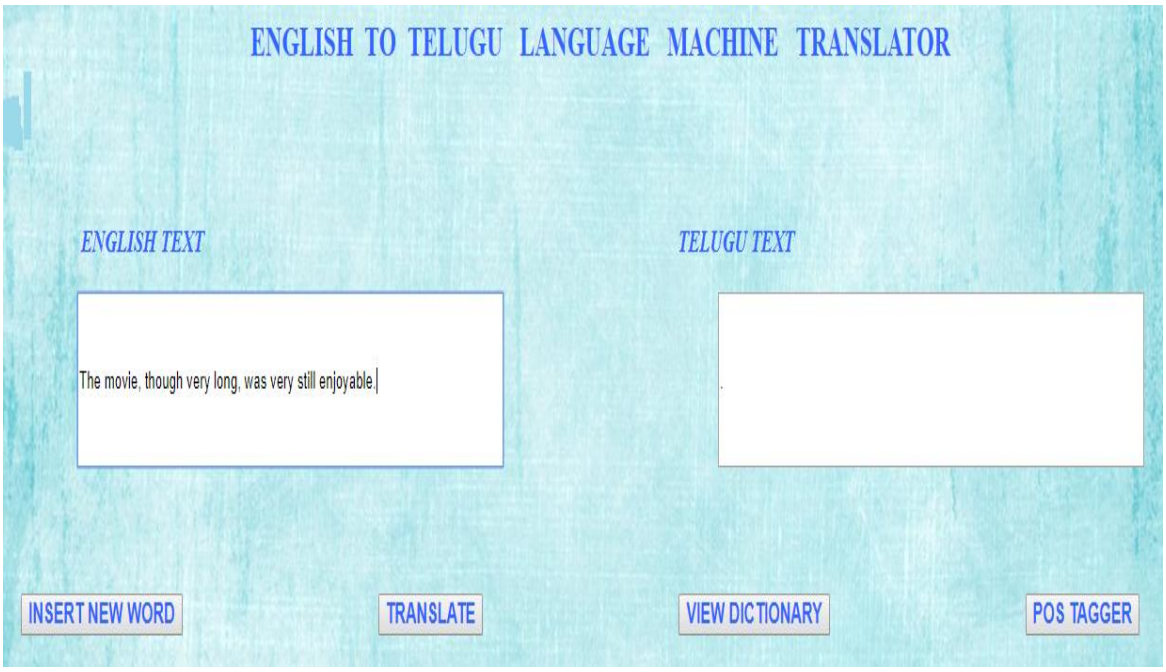


Fig 2. The Input sentence is given in the English (Source) Text Box.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

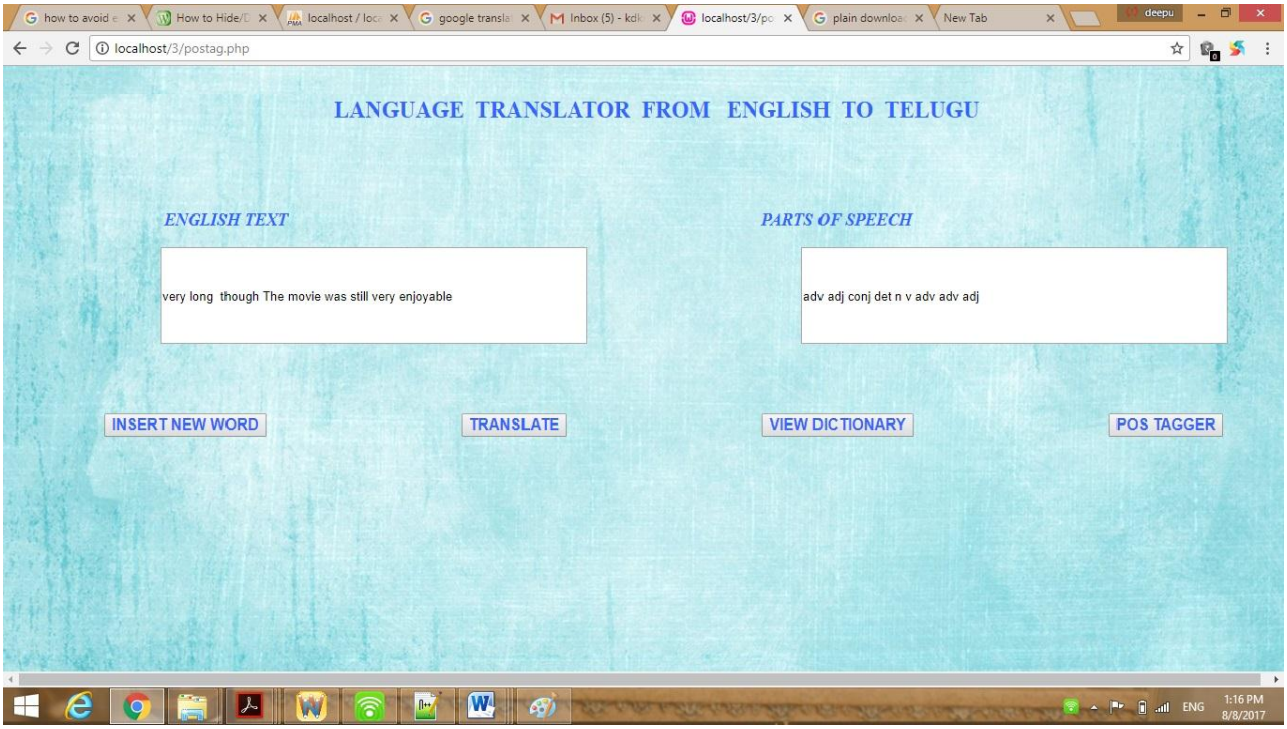


Fig 3. The POS for the Input sentence is given in the Parts Of Speech Text Box.

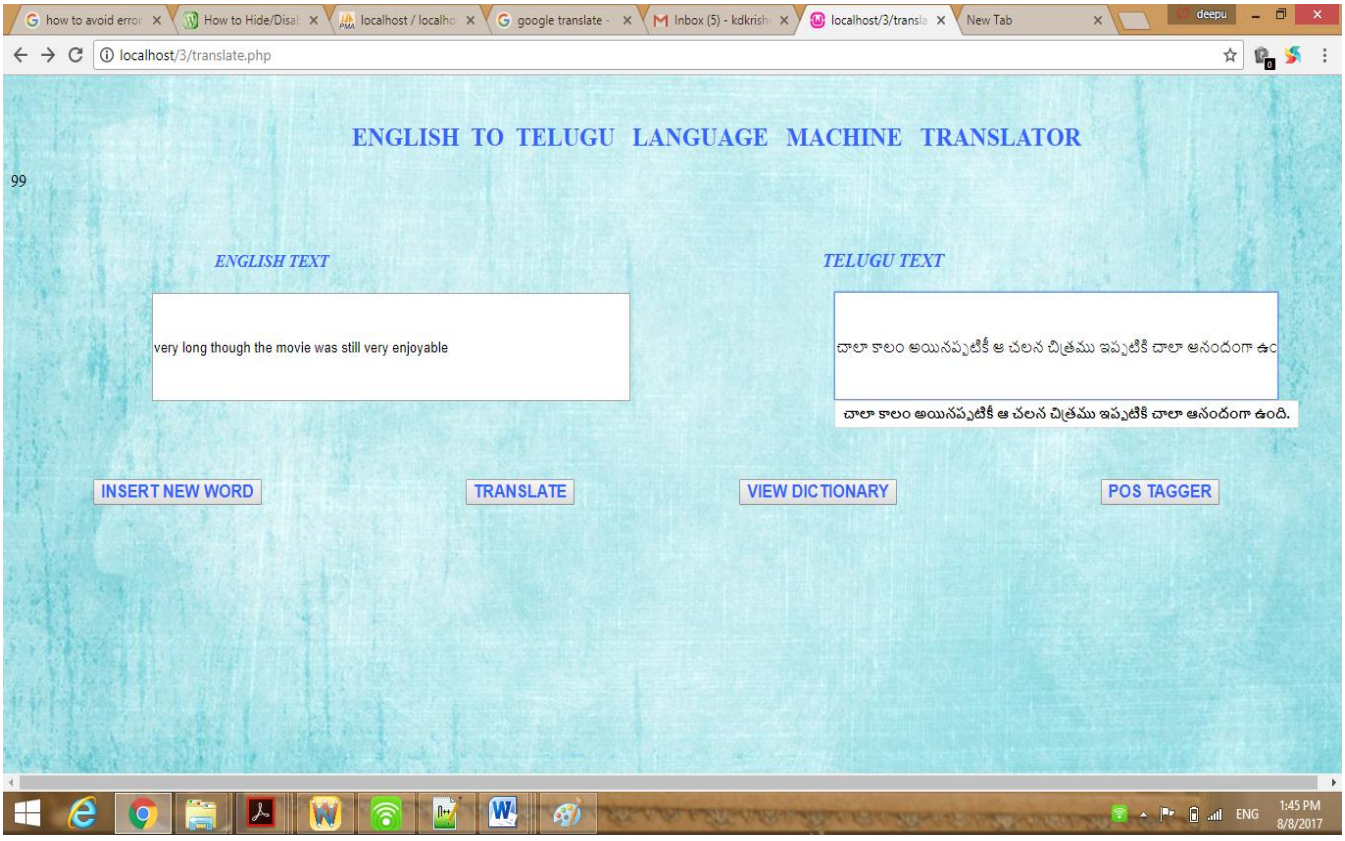
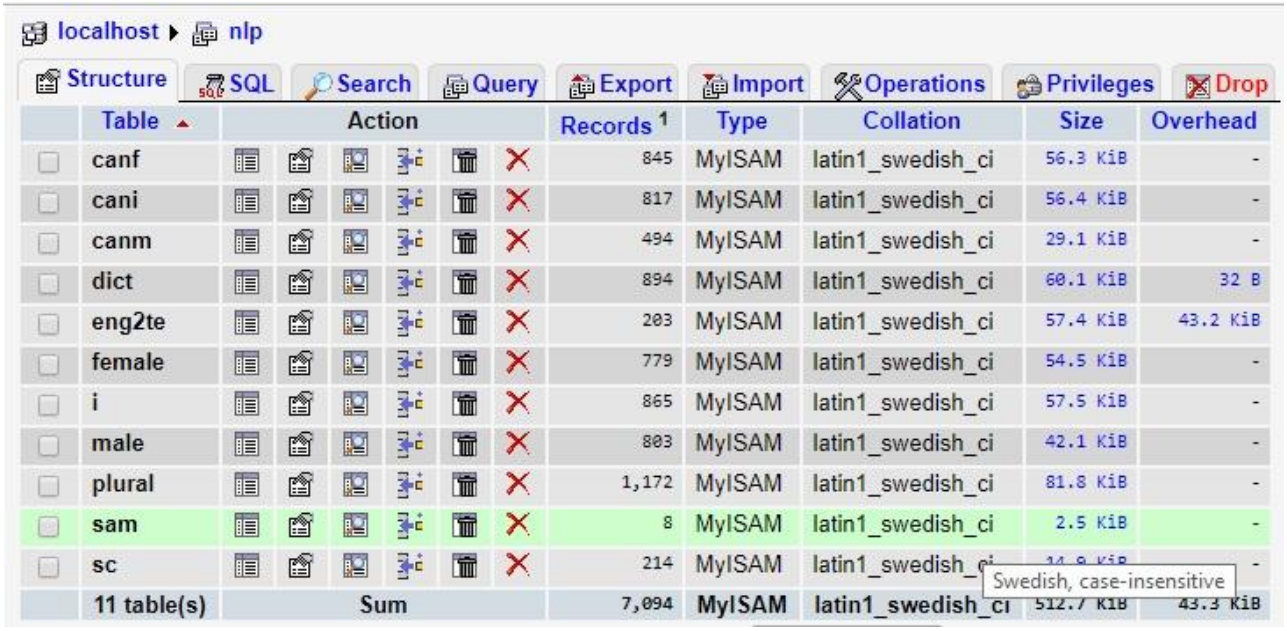


Fig 4. The Output for the Input sentence is given in Telugu (Target) Text Box.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

LIST OF SAMPLE DATABASES



localhost ▸ nlp

Structure | SQL | Search | Query | Export | Import | Operations | Privileges | Drop

	Table	Action	Records [1]	Type	Collation	Size	Overhead
☐	canf		845	MyISAM	latin1_swedish_ci	56.3 KiB	-
☐	cani		817	MyISAM	latin1_swedish_ci	56.4 KiB	-
☐	canm		494	MyISAM	latin1_swedish_ci	29.1 KiB	-
☐	dict		894	MyISAM	latin1_swedish_ci	60.1 KiB	32 B
☐	eng2te		203	MyISAM	latin1_swedish_ci	57.4 KiB	43.2 KiB
☐	female		779	MyISAM	latin1_swedish_ci	54.5 KiB	-
☐	i		865	MyISAM	latin1_swedish_ci	57.5 KiB	-
☐	male		803	MyISAM	latin1_swedish_ci	42.1 KiB	-
☐	plural		1,172	MyISAM	latin1_swedish_ci	81.8 KiB	-
☐	sam		8	MyISAM	latin1_swedish_ci	2.5 KiB	-
☐	sc		214	MyISAM	latin1_swedish_ci	14.9 KiB	-
	11 table(s)	Sum	7,094	MyISAM	latin1_swedish_ci	512.7 KiB	43.3 KiB

Swedish, case-insensitive

Fig 5. Table name: List of Tables created for Translation process

←T→	sno	eword	tword	pos
✎ ✕	1	beautiful	అందంమైన	adj
✎ ✕	2	Fruits	ఫలాలు	n
✎ ✕	3	Friends	స్నేహితులు	n
✎ ✕	4	Brothers	అన్నతమ్ములు	n
✎ ✕	5	Sisters	అక్కాచెల్లెలు	n
✎ ✕	6	are	ఉన్నాయి	v
✎ ✕	7	OEsophagus	కంఠ నాళము.	n
✎ ✕	8	Of	యొక్క	prep
✎ ✕	9	Odour	వాసన, పరిమళము	n

Fig 6. Table name: plural

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

	sno	eword	tword	pos	gender
	1	play	ఆడు	v	n
	2	went	వెళ్ళుట	v	m
	3	calling	పిలిస్తున్నాడు	v	m
	4	I	నేను	pn	n
	5	you	నీవు	pn	n
	6	but	కాని	conj	n
	7	so	అందువలన	conj	n
	8	wait	ఆగు	v	n
	9	until	వరుకు	conj	n
	10	Tomorrow	రేపు	n	n
	11	finishes	పూర్తయ్యేదాకా	v	n
	12	before	ముందు	conj	n
	13	she	ఆమె	pn	f
	14	took	తీసుకొనుట	v	n
	15	wants	కావాలి	v	n
	16	Fruits	ఫలాలు	n	p
	17	come	రా	v	n
	18	her	ఆమె	pn	f
	19	spoke	[illegible]	v	f
	20	spoke	మాట్లాడాడు	v	m
	21	speaks	మాట్లాడుతాడు	v	m
	22	speaks	మాట్లాడుతాది	v	f
	23	i'm	నేను	pn	n

Fig 8. Table name:dict

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

←T→	sno	eword	tword	pos	gender
	1	she	అమె	pn	f
	2	her	తన	pn	f
	3	called	పిలిచింది	v	f
	4	wants	కావాలి	v	n
	5	spoke	మాట్లాడింది	v	f
	6	in	లో	prep	n
	7	he	అతడు	pn	m
	8	he	అతడు	pn	m
	9	with	తో	prep	n
	10	called	పిలిచాడు	v	m
	11	and	మరియు	conj	n
	12	are	ఉన్నాయి	v	p
	13	like	ఇష్టం	v	n
	14	need	కావాలి	v	n
	15	is	ఉంది	v	n
	19	Oesophagus	కంఠ నాళము.	n	n
	17	that	అ	conj	n
	18	heard	విన్నాది	v	f
	20	Of	యొక్క	prep	n
	21	Odour	వాసన, పరిమళము	n	
	22	Odoriferousness	వాసన, పరిమళము.	n	
	23	Odorous	పరిమళించే, వాసన కొట్	adj	
	24	Odoriferous	పరిమళించే, వాసన కొట్	adj	

Fig 9. Table name: female

VIII. CONCLUSION AND FUTURE WORK

The simulation results showed that the proposed algorithm performs better with the complex sentences consisting of more than one comma's in the sentence. The proposed algorithm identifies the unlike clauses and the connective joining it. In order to perform better translation, the Subordinate Conjunction (SC) is moved to the end of the Dependent Clause thereby combining the Independent clauses together. Finally, to produce effective target translation, the both clauses with SC movement after re-ordering is merged and translated in Telugu. We can further extend our work with translating discourses of complex sentences in an effective manner.

REFERENCES

[1] Keerthi Lingam ," International Journal of Computer Applications." English to Telugu Rule based Machine Translation System: A Hybrid Approach (0975 – 8887) Volume 101– No.2, September 2014 .

[2] Sandeep Saini, "Relative clause based Text simplification for Improved English to Hindi Translation, IEEE(2015).

[3] Antony, P. J. "Machine Translation Approaches and Survey for Indian Languages." Computational Linguistics and Chinese Language Processing Vol 18 (2013): 47-78.

[4] J. Ameta, N. Joshi, I. Mathur. 2013. Improving the Quality of Gujarati-Hindi Machine Translation Through Part-of-Speech Tagging and Stemmer-Assisted Transliteration. International Journal on Natural Language Computing, Vol 3(2), pp 49-54.

[5] Hauenschild C. 1988. Discourse structure – some implications for Machine Translation, In proceedings. Of Conf. on New Directions in Machine Translation, Budapest, August 18-19 Dodrecht-Holland

[6] Crystal, D. 1988. The Cambridge Encyclopedia of Language, Cambridge Univ. Press.

[7] Suryakanthi T., Prasad S. V. A. V. and Prasad T. V., 2013 Translation of Pronominal Anaphora From English To Telugu Language, Int. J. of Adv. Computer Sc. and App., Vol. 04, No. 04

[8] Hobbs, Jerry, 1978, Resolving pronoun references, J of Lingua, April, Vol.44 pp.311-338.

[9] T.Suryakanthi and Kamlesh Sharma,"Discourse Translation from English to Telugu" ,acm.edu.in.

[10] Murphy Raymond. 2012. English grammar in use, 4th edition, Cambridge Univ. Press, New york.

BIOGRAPHY

Deepthi is a M.Tech Student in the Computer Science Department, Anil Neerukonda Institute of Technology and Sciences. She is pursuing Master of Technology (M.Tech) degree in 2015-17 from ANITS, Visakhapatnam, India. Her research interests are Machine Learning, Image Processing, Computer Networks, etc.

Keerthi, currently working as Asst Prof. in department of Computer Science and Engineering in Anil Neerukonda Institute of Technology and Sciences. She has pursued Master Of Technology (M.Tech) degree in CSE. Have overall 7 years of teaching experience. Guided many UG and PG projects as supervisor. Published several papers in international and national journals, conferences. Attended various FDP, workshops. Research areas include Image Processing, Machine Learning and Data Mining.