Volume: 7 Issue: VIII Month of publication: August 2019

DOI: http://doi.org/10.22214/ijraset.2019.8134

UNL Enconversion for Tamil Sentence

R. Baghia Laxmi[1], Dr. S. Lakshmana Pandian[2]
[1, 2]*Dept. of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India*

Abstract: Natural language processing (NLP) aims to design and build software that will analyze, understand and generate languages that humans use naturally. Machine translation is a very important application in natural language processing. Universal Networking Language (UNL) is an intermediator which exchanges the knowledge from a natural language to allow the access of its content through different languages. In this proposed work, the development of an enconverter (analyzer) process is done in order to overcome language barriers from Tamil sentence to UNL format. The grammar teaching tools like character analyzer for analyzing character, morphological analyzer and verb conjugator for the word level analysis, Parts of speech (POS) tagger, chunker and dependency parser for the sentence level analysis will be developed using machine learning based technology. The purpose is to enconvert the Tamil sentence to UNL format so that it can be easily translated to any other targeted natural languages.

Index Terms: Natural language processing (NLP), Universal Networking Language (UNL), enconverter and Tamil sentence.

I. INTRODUCTION

Natural language processing (NLP) aims to design and build software that will analyze, understand and generate languages that humans use naturally. Machine translation is a very important application in natural language processing. It is a field of computer science, artificial intelligence and linguistics and is an area of research and application that analyze how computers are used for understanding and manipulating natural language text or speech to achieve the desired tasks. NLP is concerned with the interactions between computers and human (natural) languages and it is related to the area of human-computer interaction. Major tasks in NLP involves Automatic Summarization, Coreference Resolution, Discourse Analysis, Machine Translation, Morphological Segmentation, Named Entity Recognition, Natural language generation, Natural language understanding, Parts of speech tagging, Parsing, Question answering, Relationship extraction, Sentence breaking, Sentiment analysis, Speech recognition, Information retrieval, Information extraction, Speech processing and it goes on. Challenges in natural language processing frequently involve speech recognition, natural language understanding, and natural language generation.

II. LITERARURE SURVEY

Balaji *et al.* [1] encompassed Morpho-Semantic Features for Rule-based Tamil Enconversion. The rich morphology of Tamil enables the enconversion process to be based on morpho-semantic features of the words and their preceding and succeeding context. UNL graphs are used to build a conceptual level index. N.Vignesh *et al.* [2] introduced automatic question Generator in Tamil which finds application in generating optimal questions which aids better understanding of structural and grammatical aspects of a language. This same technique can be applied to other languages only varying the basic blocks of grammar for the corresponding language. Ananthi Sheshasaaye *et al.* [3] achieved predominant role in machine translation of larger vocabulary tasks. Achieving this goal is not an easy task especially when it comes to languages like Tamil which are agglutinative in nature. Deep analysis is needed at the word level to confine the correct meaning of the word from its morphemes and categories. Language Dependent Features for UNL-Malayalam Deconversion proposed by Biji Nair *et al*. [4].This system was efficient in generating syntactically unambiguous and semantically equivalent target sentence for the UNL source sentences. Further refinement of the mapping can be done for automating the production of generation rules rather that manual rule generation. Nawab Y. Ali *et al.* [5] developed UNL-Based Machine Translation Scheme for Bangla Locative Case Constructs. It analyzed various types of Bangla locative case sentences in favor of UNL structure considering the lexicon and UNL relations they create. S. Lushanthan *et al.* [6] developed Morphological Analyzer and Generator for Tamil Language. It illustrates how the lexicon and the orthographic rules of Tamil language had been written as regular expressions using only finite state operations and how this approach had been implemented to develop a morphological generator/analyzer. Athira.K [7] proposed UNL Enconversion framework for Machine Translation which presents the building of a language independent enconversion framework of UNL. UNL framework is tested for Malayalam and English languages. In these languages the dependencies that generated between the concepts are suitable for unambiguously determining the UNL relations occurring between them. In this, enconverter handles only simple sentences.

M. F. Mridha *et al.* [8] composed Design and Implementation of an Efficient Enconverter for Bangla Language. It highlights the enconversion analyzing rules for the enconverter and indicates its usage in generating UNL expressions. The Bangla enconverter processes the given input sentence from left to right. It uses two types of windows namely, analysis window and condition window in the processing. Morphological Analyzer for Classical Tamil Text: A Rule-Based Approach developed by R. Akilan *et al.* [9] The rule based approach produce the best accuracy of tagged corpus in Classical Tamil texts, based on the 93 rules have been implemented in this analyzer. For the testing and evaluation purpose 3257 words have been taken as input of the morphological analyzer, it produces the result of 270 words are analyzed correctly 359 (11%) of words are analyzed wrongly and 194 (6%) of words are un-analyzed. Imane Taghablout *et al.* [10] designed Amazigh verb in the Universal Networking Language. It focuses on morphological analysis and semantic information of Amazigh verbs, with the aim to incorporate them into the Amazigh UNL dictionary. Translation Challenges and Universal Networking Language investigated by Baljeet Kaur Dhindsa *et al* [11].Universal Networking Language (UNL) based on Interlingua approach used especially for translation among multiple languages because it requires knowledge of UNL and of the language which user wants UNL to support. It has the advantage of saving time and money as it is faster and has high throughput as compared to that by human translator. Data Extraction from Natural Language Using UNL completed by Aloke Kumar Saha *et al.* [12]. It paves way to introduce a unique technique on data extraction – providing the user with exactly what is asked without any mimicry of unsolicited data.

III. UNIVERSAL NETWORKING LANGUAGE

The Universal Networking Language (UNL) is an electronic language in the form of a semantic network for computers to express and exchange every kind of information. This language is assumed to express meanings in the same standardized way as HTML presents its layout. The UNL represents information, i.e. meaning, sentence by sentence. Sentence information is represented as a hyper-graph having Universal Words (UWs) as nodes and relations as arcs. This hyper-graph is also represented by a set of directed binary relations, each between two of the UWs present in the sentence. The UNL expresses information classifying objectivity and subjectivity. Objectivity is expressed using UWs and relations. Subjectivity is expressed using attributes by attaching them to UWs. Nodes, or Universal Words (UWs) are words loaned from English and disambiguated by their positioning in a knowledge base (KB) of conceptual hierarchies. Function words, such as determiners and auxiliaries are represented in the form of attributes to UWs, provided that these function words contribute to the meaning.

IV. PROPOSED WORK

A process called “Enconversion” which involves various steps like Parts-of-Speech Tagging, Parts-of-Speech Parsing, Entity Identification, Relation identification, building Dictionary, Generation rules. Later on splitting, tokenizing, POS tagging, Chunking process is carried out. Enconverter generates UNL expressions from sentences (or lists of words of sentences) of a Tamil language by applying enconversion rules. The input is given as a Tamil sentence. Enconverter applies enconversion rules to the Node-list. Finally the UNL representation is produced as output. The process of converting a source language (natural language) expression into the UNL expression is referred to as enconversion. Enconversion process involves POS tagging, parsing, Entity Identification, construction of dictionary and rules to produce UNL representation.

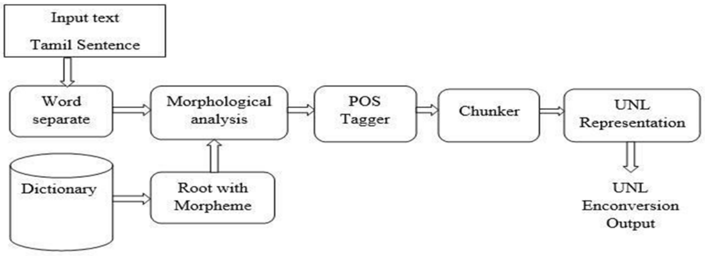


Fig 4.1 Proposed Work Architecture

POS tagging and parsing process is carried out by Stanford Parser. A natural language parser is a program that works out the grammatical structure of sentences, for instance, which groups of words go together as "phrases" and which words are the subjects or object of a verb. The Tamil sentence is given as the input text. After the formation of a text, the words can be separated. Then, the morphological analysis is done with the help of Dictionary. POS tagger and chunker are used to obtain the UNL representation. The final UNL enconversion output is obtained to easily translate any targeted natural languages. The Fig 4.1 shows the Proposed Work Architecture.

A. Morphological Analyzer

Morphology is the part of linguistics that deals with the study of words, their internal structure and partially their meanings. Morphological Analysis is the process of providing grammatical information of a word given its suffix. A morphological analyzer is a program for analyzing the morphology of an input word, it detects morphemes of any text. Morphological analysis is the process of segmenting a given word into a sequence of morphemes. The output obtained in this process is also an irregular English sentence but tense of the sentence is to be checked. It will split the sentence into word by word. After that, it will split each and every word into its corresponding root word and its morpheme components. This process can be done by the means of complex words. If suppose the word is simple, it cannot be able to split means it will directly translate into English words with the help of dictionary. The complex sentence splitter from the Tamil language can also be translated into English and based on the tense it can be translated into corresponding English words. With the help of Morphological analysis for Tamil sentence, the input can be splitter and with the Tamil to English Dictionary Tamil to English words are to be retrieved. After translation it can be able to check the tense of the input based on the tense. The corresponding English words are to be written.

A morphological analyzer is a computational tool to analyze word forms into their roots and functional elements. Tamil is one of the Dravidian languages and as such has an agglutinative grammar. It requires deep analysis at the word level to capture the correct meaning of the word from its morphemes and categories. Generally in Tamil language inflections to the root word are post propositional eventually it takes few thousand inflected form of words. Morphological analysis consists of the identification of parts of the words, or more technically, constituents of the words. This method used for analysis makes use of a stem dictionary (for identifying a valid stem), a suffix dictionary containing all possible suffixes that nouns/verbs in the language can have (to identify a valid suffix), morphotactic rules and morphophonemic rules. Morphological analyser is built in used in a way to provide a language independent framework which helps other similar languages also to work by altering the language dependent module. Tamil dictionary is collected and embedded as Btree structure.

B. Pos Tagger

The Part of speech (POS) tagging is the process of labeling a part of speech or other lexical class marker to each and every word in a sentence. It is similar to the process of tokenization for computer languages. Labelling words for POS can be done by dictionary lookup and/or some sort of process. Identifying POS can be seen as a prerequisite to parsing, and/or a result of morphological analysis in its own right.POS tagging is considered as an important process in speech recognition, natural language parsing, information retrieval and machine translation. Tamil being a Dravidian language has a very rich morphological structure which is agglutinative. Tamil words are made up of lexical roots followed by one or more affixes. So tagging a word in a language like Tamil is very complex. The main challenges in Tamil POS tagging are solving the complexity and ambiguity of words. For example, the words are assigned in the grammatical category

கோயிலில்	ஆறு	அடி	உயரமான	மணி	உள்ளது .
NN	CRD	NN	ADJ	NN	VF

C. Chunker

A subsequent step after tagging focuses on the identification of basic structural relations between groups of words. This is usually referred to as phrase chunking.

a) Input: Word sequence and POS tags
b) Output: A single best Chunk Tag for each word along with its POS tag.
1) A “chunk” is a continuous non-overlapping sequence of words
2) Chunker finds such sequences, often using tagged text as input
3) Chunk rules can be as simple as regular expressions
4) Chunkers can allow embedding, but typically only to a shallow level
5) Tamil being an agglutinative language have a complex morphological and syntactical structure.
6) It is a relatively free word order language but in the phrasal and clausal construction it behaves like a fixed word order language.

For example, Chunk Tags are assigned. The IOB tags are used to indicate the boundaries for each chunk

a) B – the current word is the beginning of a chunk, which may be followed by another chunk.

b) O - indicates the boundary of the sentence.

c) I – the current word is inside a chunk**.**

B-NP B-NP I-NP B-NP I-NP B-VP
கோயிலில் ஆறு அடி உயரமான மணி உள்ளது .

D. Tamil to English Dictionary

By using Tamil to English Dictionary the Chunked Tamil sentence is translated to English sentence and then UNL representation is done using the UNL Relations and UNL Attributes. Transliteration is the process of transferring a word from the alphabet of one language to another. Transliteration helps people pronounce words and names in foreign languages. A translation tells the meaning of words in another language. For example, Tamil sentence is translated to English sentence.

கோயிலில் ஆறு அடி உயரமான மணி உள்ளது .

Six feet tall bell is in the temple.

E. UNL Representation

UNL vocabulary consists of Universal Words, Relations and Attributes.

1) Universal Words (UWs): Labels that represent word meanings.

2) Relation Labels: Tags that represent the relationship between Universal Words.

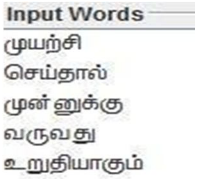
3) Attribute Labels: represent the further definition or additional information, which appears in the sentence.

Binary relations are the building blocks of UNL expressions. They are made up of a relation and two UWs. There are two forms for expressing the UNL expressions, one is the table form and the other is the list form. The table form of a UNL expression is more readable than the list form, but the list form of a UNL expression is more compact than the table form. Any component, such as a word, phrase or title and a sentence of a natural language can be represented with UNL expressions. A UNL expression therefore consists of a UW or a set of binary relations. In UNL documents, a UNL expression for a sentence is enclosed by the tags {unl} and {/unl}.

Example : Six feet tall bell is in the temple.
[S]
[UNL]
[W]
Six (icl>qua).@generic:0 feet(icl>ben).@generic:1 tall(icl>man). @generic:2
bell(icl>ins). @generic:3
temple(icl>plc>.@generic:4

After the analysing process, the words get separated. The separation of input words are shown in Fig 5.3.The splitting of words taken place in the output. It splits the verb, tense, etc., and it also shows the working of morphological Analyzer. The Fig 5.4 shows the different types of tags.Tags used are English_abbreviated, English_expanded, Tamil_abbreviated and Tamil_expanded. It also use the dictionary. The different types of tags are marked in this snapshot.

1 man 2
2 ins 3
3 plc 4
[/R]
[/UNL]

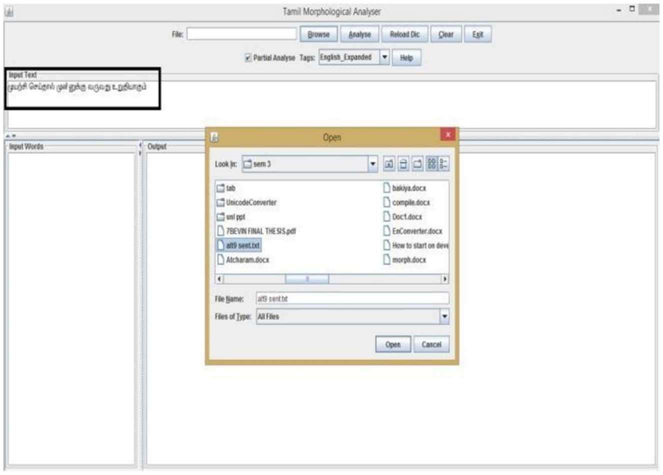


Input Words
முயற்சி
செய்தால்
முன் னுக்கு
வருவது
உறுதியாகும்

Here *qua* (quantity), *ben (*beneficiary*), man (*manner), *ins* (instrument) and *plc* (place) are the UNL relations. These UWs have restrictions mentioned in parentheses for the purpose of denoting a unique sense.

V. EXPERIMENTAL RESULTS

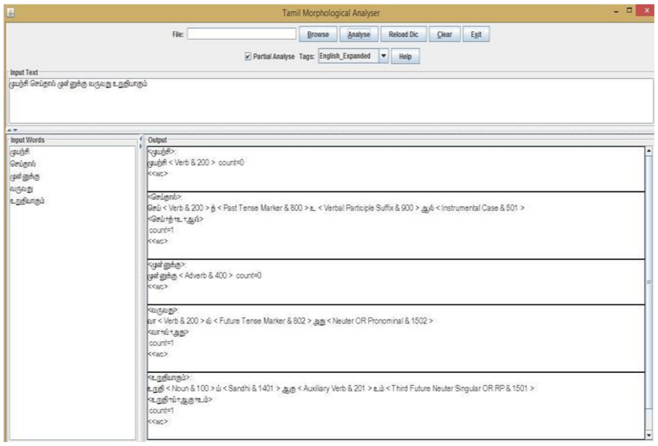
The Fig 5.1 shows the input. The input text is given in type of a box. The output forms after the input text is given. The Fig 5.2 forms the output.



Input Text
முயற்சி செய்தால் முன்னுக்கு வருவது உறுதியாகும்

```
tab                      bakiya.docx
UnicodeConverter         compile.docx
uni ppt                  Doc1.docx
7BEVIN FINAL THESIS.pdf  EnConverter.docx
alt9 sent.txt            How to start on deve
Atcharam.docx            morph.docx

File Name:    alt9 sent.txt
Files of Type: All Files
```

Fig 5.1 Iutput

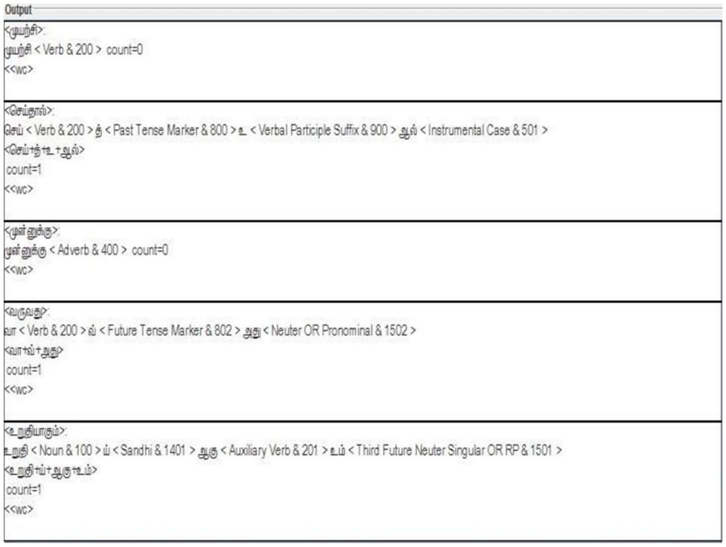


Input Text
முயற்சி செய்தால் முன்னுக்கு வருவது உறுதியாகும்

Input Words
முயற்சி
செய்தால்
முன்னுக்கு
வருவது
உறுதியாகும்

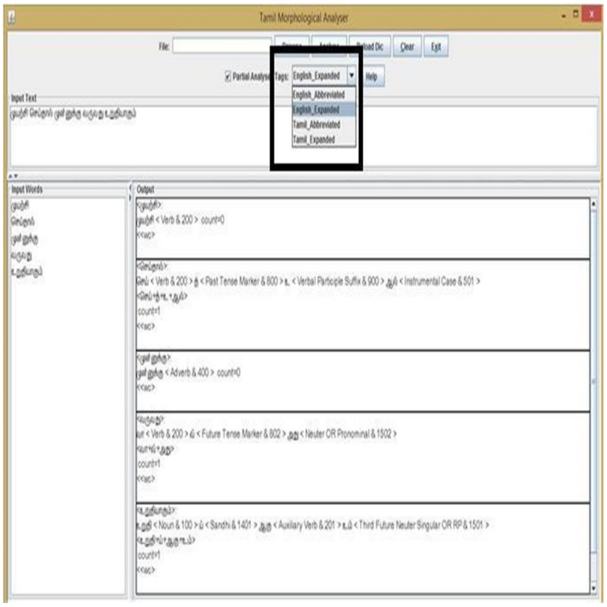
```
Output
<முயற்சி>:
முயற்சி < Verb & 200 >  count=0
<<wc>

<செய்தால்>:
செய் < Verb & 200 > த் < Past Tense Marker & 800 > உ < Verbal Participle Suffix & 900 > ஆல் < Instrumental Case & 501 >
<செய்+த்+உ+ஆல்>
 count=1
<<wc>

<முன்னுக்கு>:
முன்னுக்கு < Adverb & 400 >  count=0
<<wc>

<வருவது>:
வா < Verb & 200 > வ் < Future Tense Marker & 802 > அது < Neuter OR Pronominal & 1502 >
<வா+வ்+அது>
 count=1
<<wc>

<உறுதியாகும்>:
உறுதி < Noun & 100 > ய் < Sandhi & 1401 > ஆகு < Auxiliary Verb & 201 > உம் < Third Future Neuter Singular OR RP & 1501 >
<உறுதி+ய்+ஆகு+உம்>
 count=1
<<wc>
```

Fig 5.2 Output



```
Output
<முயற்சி>:
முயற்சி < Verb & 200 >  count=0
<<wc>

<செய்தால்>:
செய் < Verb & 200 > த் < Past Tense Marker & 800 > உ < Verbal Participle Suffix & 900 > ஆல் < Instrumental Case & 501 >
<செய்+த்+உ+ஆல்>
 count=1
<<wc>

<முன்னுக்கு>:
முன்னுக்கு < Adverb & 400 >  count=0
<<wc>

<வருவது>:
வா < Verb & 200 > வ் < Future Tense Marker & 802 > அது < Neuter OR Pronominal & 1502 >
<வா+வ்+அது>
 count=1
<<wc>

<உறுதியாகும்>:
உறுதி < Noun & 100 > ய் < Sandhi & 1401 > ஆகு < Auxiliary Verb & 201 > உம் < Third Future Neuter Singular OR RP & 1501 >
<உறுதி+ய்+ஆகு+உம்>
 count=1
<<wc>
```

Fig 5.3 Word separate & Analysing Process



Input Text

முயற்சி செய்தால் முன்னுக்கு வருவது உறுதிப்பாகும்

Input Words

முயற்சி
செய்தால்
முன்னுக்கு
வருவது
உறுதிப்பாகும்

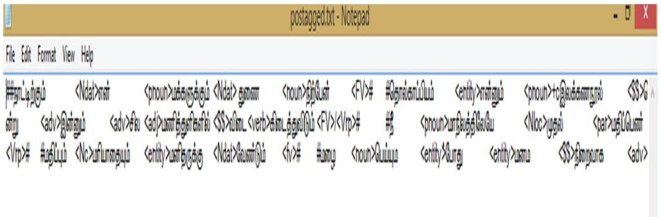
Output

```
<முயற்சி>
முயற்சி < Verb & 200 > count=0
<<wc>

<செய்தால்>
செய் < Verb & 200 > த் < Past Tense Marker & 800 > உ < Verbal Participle Suffix & 900 > ஆல் < Instrumental Case & 501 >
<செய்+த்+உ+ஆல்>
count=1
<<wc>

<முன்னுக்கு>
முன்னுக்கு < Adverb & 400 > count=0
<<wc>

<வருவது>
வா < Verb & 200 > வ் < Future Tense Marker & 802 > அது < Neuter OR Pronominal & 1502 >
<வா+வ்+அது>
count=1
<<wc>

<உறுதிப்பாகும்>
உறுதி < Noun & 100 > ப் < Sandhi & 1401 > ஆகு < Auxilary Verb & 201 > உம் < Third Future Neuter Singular OR RP & 1501 >
<உறுதி+ப்+ஆகு+உம்>
count=1
<<wc>
```

Fig 5.4 Different Types of Tags

Part-of-speech (POS) tagging, also called grammatical tagging, is the process of assigning POS tags to each and every word in a sentence.It is like assigning the grammatical category such as Noun, Verb, Adjective, Adverb etc . It is shown in Fig 5.5

1) *Input:* a string of words (sentence)
2) *Output:* a single best tag for each word (POS Tagged sentence)

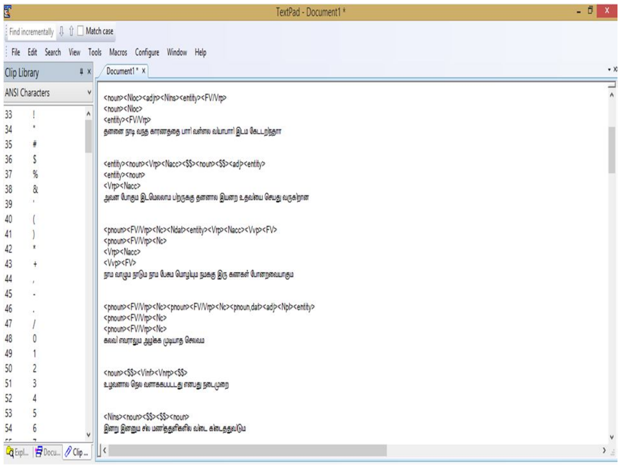


```
##நாட்டிற்கும்	<Ndat>என்	<pnoun>மக்களுக்கும் <Ndat> துணை	<noun>நிற்பேன்	<FV>#	#தொல்காப்பியம்	<entity>என்னும்	<pnoun>+தொல்காப்பியர்	<SS>G
என்று	<adv>இன்னும்	<adv>சில <adj>பணித்துளிகளில் <SS>விடை <verb>கிடைத்துவிடும் <FV>/<Vrp>#	#நீ	<pnoun>மாநிலத்திலேயே	<Noc>முதல்	<par>மதிப்பெண்
<Vrp>#	#மதிப்பும்	<Nc>மரியாதையும்	<entity>மனிதருக்கு	<Ndat>வேண்டும்	<fv>#	#மழை	<noun>பெய்யும்	<entity>போது	<entity>மனம்	<SS>நிறைவாக	<adv>
```

Fig 5.5 POS

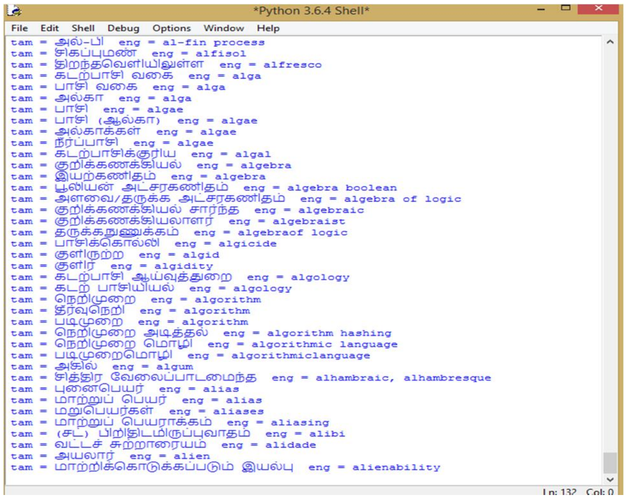
A subsequent step after tagging focuses on the identification of basic structural relations between groups of words. This is usually referred to as phrase chunking. It is shown in Fig 5.6

1) *Input:* Word sequence and POS tags
2) *Output:* A single best Chunk Tag for each word along with its POS tag.



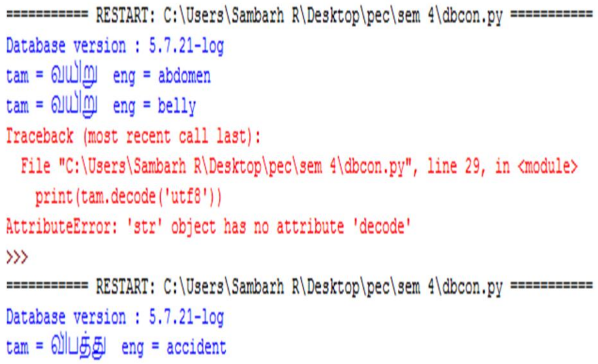
```
<noun><Nloc><adjn><Nins><entity><FV/Vrp>
<noun><Nloc>
<entity><FV/Vrp>
தன்னை நாடி வந்த காரணத்தை பாரி வள்ளல் வியாபாரி இடம் கேட்டறிந்தார்


<entity><noun><Vrp><Nacc><SS><noun><SS><adj><entity>
<entity><noun>
<Vrp><Nacc>
அவன் போகும் இடமெல்லாம் பிறருக்கு தன்னால் இயன்ற உதவியை செய்து வருகிறான்


<pnoun><FV/Vrp><Nc><Ndat><entity><Vrp><Nacc><Vvp><FV>
<pnoun><FV/Vrp><Nc>
<Vrp><Nacc>
<Vvp><FV>
நாம் வாழும் நாடும் நாம் பேசும் மொழியும் நமக்கு இரு கண்கள் போன்றவையாகும்


<pnoun><FV/Vrp><Nc><pnoun><FV/Vrp><Nc><pnoun,dat><adj><Npl><entity>
<pnoun><FV/Vrp><Nc>
<pnoun><FV/Vrp><Nc>
கல்வி எவராலும் அழிக்க முடியாத செல்வம்


<noun><SS><Vinf><Vnrp><SS>
உழவனால் நெல் விளைவிக்கப்பட்டது என்பது நடைமுறை


<Nins><noun><SS><SS><noun>
இன்று இன்னும் சில மணித்துளிகளில் விடை கிடைத்துவிடும்
```

Fig 5.6 Chunking



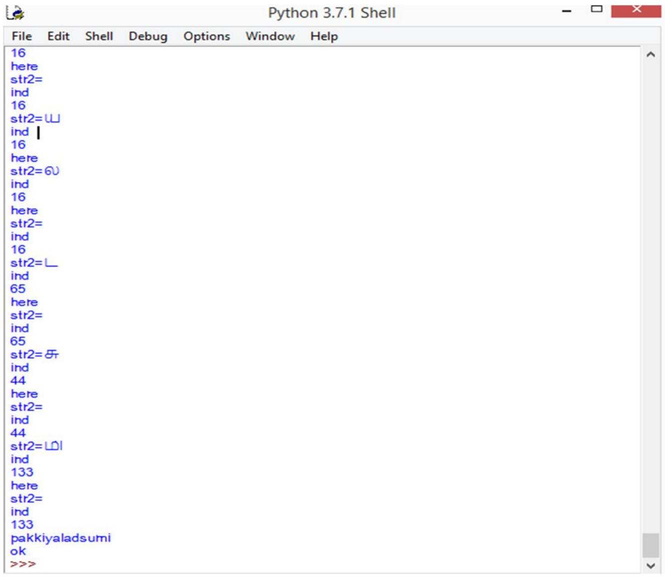
```
tam = அல்-பி  eng = al-fin process
tam = சிகப்புமண்  eng = alfisol
tam = திறந்தவெளியிலுள்ள  eng = alfresco
tam = கடற்பாசி வகை  eng = alga
tam = பாசி வகை  eng = alga
tam = அல்கா  eng = alga
tam = பாசி  eng = algae
tam = பாசி (ஆல்கா)  eng = algae
tam = அல்காக்கள்  eng = algae
tam = நீர்ப்பாசி  eng = algae
tam = கடற்பாசிக்குரிய  eng = algal
tam = குறிக்கணக்கியல்  eng = algebra
tam = இயற்கணிதம்  eng = algebra
tam = பூலியன் அட்சரகணிதம்  eng = algebra boolean
tam = அளவை/தருக்க அட்சரகணிதம்  eng = algebra of logic
tam = குறிக்கணக்கியல் சார்ந்த  eng = algebraic
tam = குறிக்கணக்கியலாளர்  eng = algebraist
tam = தருக்கநுணுக்கம்  eng = algebraof logic
tam = பாசிக்கொல்லி  eng = algicide
tam = குளிருற்ற  eng = algid
tam = குளிர்  eng = algidity
tam = கடற்பாசி ஆய்வுத்துறை  eng = algology
tam = கடற் பாசியியல்  eng = algology
tam = நெறிமுறை  eng = algorithm
tam = தீர்வுநெறி  eng = algorithm
tam = படிமுறை  eng = algorithm
tam = நெறிமுறை அடித்தல்  eng = algorithm hashing
tam = நெறிமுறை மொழி  eng = algorithmic language
tam = படிமுறைமொழி  eng = algorithmiclanguage
tam = அகில்  eng = algum
tam = சித்திர வேலைப்பாடமைந்த  eng = alhambraic, alhambresque
tam = புனைபெயர்  eng = alias
tam = மாற்றுப் பெயர்  eng = alias
tam = மறுபெயர்கள்  eng = aliases
tam = மாற்றுப் பெயராக்கம்  eng = aliasing
tam = (சட்) பிறிதிடமிருப்புவாதம்  eng = alibi
tam = வட்டச் சுற்றாரையம்  eng = alidade
tam = அயலார்  eng = alien
tam = மாற்றிக்கொடுக்கப்படும் இயல்பு  eng = alienability
```

Fig 5.7 Dictionary Database



```
=========== RESTART: C:\Users\Sambarh R\Desktop\pec\sem 4\dbcon.py ===========
Database version : 5.7.21-log
tam = வயிறு  eng = abdomen
tam = வயிறு  eng = belly
Traceback (most recent call last):
  File "C:\Users\Sambarh R\Desktop\pec\sem 4\dbcon.py", line 29, in <module>
    print(tam.decode('utf8'))
AttributeError: 'str' object has no attribute 'decode'
>>>
=========== RESTART: C:\Users\Sambarh R\Desktop\pec\sem 4\dbcon.py ===========
Database version : 5.7.21-log
tam = விபத்து  eng = accident
```

Fig 5.8 Translation



```
16
here
str2=
ind
16
str2=ய
ind
16
here
str2=ல
ind
16
here
str2=
ind
16
str2=ட
ind
65
here
str2=
ind
65
str2=சு
ind
44
here
str2=
ind
44
str2=மி
ind
133
here
str2=
ind
133
pakkiyaladsumi
ok
>>>
```

Fig 5.9 Transliteration

The Fig 5.7 shows the dictionary database which converts Tamil to English dictionary. The Fig 5.8 shows the translation with single Tamil word to different English words. The Fig 5.9 shows the transliteration.

VI. CONCLUSION AND FUTURE IMPROVEMENT

A language independent framework for UNL has been developed, implemented and tested. A tool for Morphological Analyzer is created. Thus, the proposed work has been completed by using enconversion and morphological analysis with the help of universal networking language. Firstly, the input is given as a Tamil sentence. After the splitting of words, the morphological analysis is done.

The future work can be further processed by parsing, summarization and also done using different rich techniques. Since the language is morphological rich and are agglutinative in nature building as an accurate system for the Tamil language is a challenging task. Challenges in natural language processing frequently involve speech recognition, natural language understanding, and natural language generation. Finally, the formation of enconversion output can be easily translated to any other targeted languages.

REFERENCES

[1] J.Balaji ,T. V. Geetha , Ranjani Parthasarathi and Madhan Karky,“Morpho-Semantic Features for Rule- based Tamil Enconversion” , International Journal of Computer Applications (IJCA), Vol. No. 26, Issue No.6, pp no.11-18, July 2011.

[2] N.Vignesh and S. Sowmya, “Automatic question Generator in Tamil”, International journal of Engineering Research & technology(IJERT),Vol. No.2, Issue No.10,October 2013.

[3] Ananthi Sheshasaayee and Angela Deepa. V.R, “The Role of Morphological Analyzer and generator for Tamil language in Machine Translation Systems”, International Journal of Computer Science and Engineering (IJCSE), Vol. No.2, Issue No.5, pp no.107-111, 2014.

[4] Biji Nair, Rajeev R and Elizabeth Sherly, “Language Dependent Features for UNL-Malayalam Deconversion”, International Journal of Computer Applications (IJCA),Vol. No.100, Issue No.6,pp no.37-41, August 2014.

[5] Nawab Y. Ali, Golam .S and Ameer.A ,“UNL-Based Machine Translation Scheme for Bangla Locative Case Constructs”, International Journal of Information and Education Technology(IJIET), Vol. No.4, Issue No. 5, pp no. 454-458, October 2014.

[6] S. Lushanthan, A. R. Weerasinghe and D. L. Herath,“Morphological Analyzer and Generator for Tamil Language” , IEEE International Conference on Advances in ICT for Emerging Regions (ICTer),Vol No.7, Issue No.5, pp no.190–196, December 2014.

[7] Athira.K, “UNL Enconversion framework for Machine Translation”,International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),Vol. No. 4, Issue No. 4,pp no.1412-1417, April 2015.

[8] M. F. Mridha, Aloke Kumar Saha, Md. Akhtaruzzaman Adnan, MollaRashied Hussein and Jugal Krishna Das, “Design and Implementation of an Efficient Enconverter for Bangla Language”, ARPN Journal of Engineering and Applied Sciences(ARPN),Vol. No. 10, Issue No. 15, pp no. 6543-6548, August 2015.

[9] M. F. Mridha, Aloke Kumar Saha, Md. Akhtaruzzaman Adnan, Molla Rashied Hussein and Jugal Krishna Das, “Design and Implementation of an Efficient Enconverter for Bangla Language”, ARPN Journal of Engineering and Applied Sciences(ARPN),Vol. No. 10, Issue No. 15, pp no. 6543-6548, August 2015.

[10] R. Akilan and E. R.Naganathan, “Morphological Analyzer for Classical Tamil Text: A Rule-Based Approach”, ARPN Journal of Engineering and Applied Sciences(ARPN), Vol. No. 10, Issue No. 20, pp no. 9325-9330,November 2015.

[11] Imane Taghablout, FadouaAtaa Allah and Mohamed marraki, “Amazigh verb in the Universal Networking Language”, IEEE Conferences of Computer Systems and Applications (AICCSA),Vol. No.4, Issue No.6,pp no. 1-4, November 2015.

[12] Baljeet Kaur Dhindsa and Dharam Veer Sharma, “Translation Challenges and Universal Networking Language”, International Journal of Computer Applications(IJCA),Vol. No. 133, Issue No.15,pp no. 36-40,January2016.

[13] Aloke Kumar Saha, M. F. Mridha, Jahir IbnaRafiq and Jugal Krishna Das, “Data Extraction from Natural Language Using Universal Networking Language”, IEEE International Conference on Current Trends in Computer, Electrical, Electronics and Communication" (ICCTCEEC) ,Vol No.2, Issue No. 7,pp no.24-29, September 2017.

[14] Hiroshi Uchida, Meiying Zhu, “The Universal Networking Language beyond Machine Translation”, UNDL Foundation, September 2009

[15] Md. Ershadul H. Choudhury, Nawab Yousuf Ali, Mohammad Zakir Hussain Sarkar, Md. Ahsan Razib,“Bridging Bangla to Universal Networking Language- A Human Language Neutral Meta- Language”,December2004.

[16] Bhattacharyya, “Multilingual Information Processing Using Universal Networking Language”, Indo UK Workshop on Language Engineering for South Asian Languages (LESAL), April 2001.

[17] Md. Nawab Yousuf Ali, Jugal Krishna Das , S. M. Abdullah Al-Mamun and Md Ershadul H.Choudhury, “Specific Features of a Converter of Web Documents from Bengali to Universal Networking Language”,IEEE International Conference on Computer and Communication Engineering,Vol No. 8,Issue No. 2, pp no. 726 - 731, 2008.

[18] Md. Nawab Yousuf Ali, Abu Mohammad Nurannabi,M. Ameer Ali, Jugal Krishna Das and Golum Farook Ahmed, “Conversion of Bangla Sentence for Universal Networking Language”, IEEE International Conference on Computer and Information Technology (ICCIT), Vol No.6, Issue No. 2, pp no. 108-113, December 2010.

[19] AjiNugraha , SantosaKasmaji and AyuPurwarianti, “Employing Natural Language Processing to Analyse Grammatical Error in a Simple Japanese Sentence”, IEEE International Conference on Electrical Engineering and Informatics(ICEEI), Vol No. 3, Issue No. 7, pp no. 82-86, August 2015.

[20] Seema Shukla and Dr. Usha Sinha , “Categorizing Sentence Structures for Phrase Level Morphological Analyzer for English to Hindi RBMT”, IEEE International Conference on Cognitive Computing and Information Processing(CCIP),Vol No. 4, Issue No. 6,pp no. 1-6, 2015.

[21] Rajeswari Sridhar, Pavithra Sethuraman and KashyapKrishnakumar, “English to Tamil machine translation system using universal networking language”, March 2016.

[22] For Universal Networking Language: Universal Networking Digital Language Foundation. http://www.undl.org/