

# A HYBRID MODEL FOR PHRASE CHUNKING EMPLOYING ARTIFICIAL IMMUNITY SYSTEM AND RULE BASED METHODS

Bindu.M.S<sup>1</sup> and Sumam Mary Idicula<sup>2</sup>

<sup>1</sup>Department of Computer Science, Mahatma Gandhi University, Kottayam, India

bindu\_aisoo@rediffmail.com

<sup>2</sup>Dept. of Computer Science, CUSAT, Cochi, India

sumam@cusat.ac.in

## ABSTRACT

*Natural language Understanding (NLU), an important field of Artificial Intelligence (AI) is concerned with the speech and language understanding between human and computer. Understanding language means knowing what concept a word or phrase stands for and how to link them to form meaningful sentence. Identification of phrases or phrase chunking is an important step in natural language understanding (NLU). Chunker identifies and divides sentences into syntactically correlated word groups. Question Answering (QA) systems, another important application of Artificial Intelligence (AI) mostly requires retrieval of nouns or noun phrases as answers to the questions raised by the users. Also Chunking is an important preprocessing step in full parsing. Due to high ambiguity of natural language, exact parsing of text may become very complex. This ambiguity may be partially resolved by using chunking as an intermediate step. To the best of our knowledge no known work or tag set is available for phrase chunking in Malayalam. To separate the chunks in a document it must be labeled with parts-of-speech (POS) tags. POS Tagging is a difficult task in Malayalam as it is a complex and compounding language. In this paper we describe the application of artificial immunity system (AIS) for chunking which is implemented and obtained an accurate output with 96% precision and 93% recall. This system is tested on corpuses collected from reputed news papers and magazines. These corpuses contained documents from five different domains such as sports, health, agriculture, science and politics and each document contained sentences –simple, compound, complex-of various levels of complexity. POS tag set with 52 tags is developed for preparing the tagged corpus for Malayalam. The phrase tag set contains 20 phrase tags.*

## KEYWORDS

*Human immune system, Self Test, POS tagger, Detector set, Phrase tags*

## 1. INTRODUCTION

Language is used for communication. To have communication between human and computer a program is to be build to understand the spoken language. NLU requires extensive knowledge of the language such as parts-of-speech (POS), Phrase structures, the way to link theses phrases to form sentences etc [1]. NLU systems need analysis and generation of phrases in natural language.

In Information Retrieval and NLP, QA is the task of automatically answering a question posed in natural language. To find the answer, questions are analyzed and determine the type of expected answers. Then the answer retrieval module retrieves either a document, passage or a phrase to answer the query. Questions such as ‘when’, ‘where’, ‘what’, ‘why’ etc. mostly requires an answer that contains a noun or a noun phrase [2].

Full parsing is the syntactic analysis of a text, made of tokens to determine the grammatical structure with respect to a given formal grammar. By this analysis it will be clear how words are combined to get phrases and phrases into sentences [3]. But parsing in human languages is not

so easy compared to the parsing in interpreter or compiler, because there is substantial ambiguity with the human language structure. Hence natural language parsing is carried out in different stages. Full parsing is an extremely difficult process for morphologically rich language like Malayalam. Hence an intermediate step is essential. Abney has proposed text chunking as a useful step for full parsing since it lays foundation for further levels of analysis [4].

Full parsing is expensive and is not very robust. But chunking is more robust, efficient, faster and sufficient for many applications. Time required to parse a sentence grows with the cube of the length of the sentence multiplied by factor  $O(G)^2$  where  $G$  is the size of the grammar [5].

Abney (1994) describes chunking as a natural phenomenon in the following words.

“(when I read) (a sentence) (I read it) (a chunk) (at a time)”

Chunking is a shallow parsing technique also called light parsing for extracting nonoverlapping segments from a stream of data. Chunker can be consecutive or interposed. A typical chunk consists of a single context word surrounded by a constellation of function words and each chunk contains a head word. Some words in a sentence may not be grouped into a chunk. A phrase chunker identifies phrases like verb phrase, noun phrase etc. in a sentence but does not either specify their internal structure or their role in the main sentence [6].

Rule based and statistical methods are normally adopted for the implementation of phrase Chunker. Rule based method require manual preparation of rules which is cumbersome for Malayalam language since there are many variations possible in the structure. Statistical methods are developed using annotated corpora which is not available for our purpose. Hence we have first selected corpus from various magazines and newspapers and added POS tags.

Malayalam belongs to the Dravidian family of languages and is one of the 4 major languages of this family. It is one of the 22 scheduled languages of India with official language status in the state of Kerala. It is spoken by 35.9 million people [7]. Malayalam is a morphologically rich agglutinative language and relatively of free order. Also Malayalam has a productive morphology that allows the creation of complex words which are often highly ambiguous. Hence POS tagging involves solving the complexity and ambiguity of the words [8]. Many words that occur in Malayalam texts are compound words which are not listed in any catalogue or lexicon. A large percentage of words also show ambiguity regarding lexical category. In many other languages, POS taggers have used tag set derived from Penn Tree Bank or Brown corpus. But this is not enough for our Malayalam POS tagger developed for the purpose of information Retrieval. We needed a tag set which also consider semantics in its development.

Phrase chunker is implemented using a novel approach based on Artificial Immunity system. AIS is a new branch of AI used for intelligent problem solving technique in optimization and scheduling using the theory of human Immunity system. Human Immunity system is an efficient defence mechanism which constitutes two layers of defence-Innate Immune system and Adaptive Immune system. Innate immunity available since birth is accomplished by providing different types of barriers to the entry of the foreign agents into our body. Skin and internal mucosal surfaces form an initial barrier to all infections. Innate Immunity is non-specific to pathogens. Adaptive Immune system is pathogen specific. Whenever there is an attack from a specific microbe this immune system is activated and antibodies are produced. Adaptive Immune system is governed by Lymphocytes comprised of B-cells (generated in the bone marrow) and T-cells (generated in the Thymus) play a significant role in recognition and elimination of pathogens [9].

AIS algorithms are more efficient than the classical heuristic scheduling algorithms. Also they are more successful than Genetic algorithms [10]. There are three algorithms widely applied: clonal selection algorithms, immune network algorithms and negative selection algorithms.

## **2. RELATED WORKS**

An artificial immunity system principle is explored in intelligent robotics system. They have implemented a behaviour arbitration mechanism for the robots to choose the best option when an abnormal situation arises [11]. Paper [12] explains the use of negative selection algorithm for the classification of EEG signals and proved that this method reveals very promising performance. The general purpose algorithm based on the clonal selection and affinity maturation process in an adaptive immunity system is presented in paper [13] which is capable of solving complex engineering tasks like multimodal and combinatorial optimization. Clonal selection algorithms are applied for edge detection problems in pattern recognition and computer vision. The principle of hyper mutation and receptor editing are exploited for the above purpose [14]. Akshat Kumar and Shivashankar B Nair described the working of English Grammar checking system using adaptive immunity system [15].

In Hindi, Tamil, Telugu, Bengali languages several works were reported but they were mainly using statistical approaches like conditional random field, hidden markov model and maximum entropy markov models [16] [17] [18]. In Tamil few taggers are developed employing support vector machines (SVM) and they obtained a tagging accuracy of 95.82%.

Many researchers have used various machine learning methods and their combinations for chunking. Noun phrase chunking is an important and useful task in many natural language applications and it is studied well in languages such as English and French. But for under sourced language like Malayalam Corpus based NLP tasks are at dead lock due to the unavailability of lexicons and taggers.

AIS theory is not yet used for the development of phrase chunker. In this paper an AIS based phrase chunker is presented. It is easy and possible to make any modifications to an AIS based system as this system works on the principle of generation and deletion of self/non self patterns.

## **3. OVER VIEW OF MALAYALAM LANGUAGE**

In Malayalam language internal structures of chunks often have unrestricted order; hence chunk identification becomes a challenging task in comparison to English text. Malayalam sentence is a sequence of words where words may appear in any order and each word can be a combination of any number of stems and affixes. Even though there is no specific order for the words in the sentence, within a chunk word categories are related.

### **3.1 Basic Word Types**

#### **3.1.1 Nouns**

Noun is classified into material noun and abstract noun. The subclasses of material nouns are proper nouns, common nouns, pronouns and collective nouns. Abstract noun is further classified into quality nouns and verbal nouns. Pronouns are available in twelve different forms [19].

#### **3.1.2 Verbs**

Verbs are divided into four categories based on their meaning, behaviour, feature and importance. The first type is divided into transitive and intransitive. Another classification based on the behaviour is simple verbs and causatives. Third type is classified into strong and weak verbs. Last division is according to its importance and is named as finite and infinite verbs.

### **3.1.3 Qualifiers**

Three types of qualifiers are there in Malayalam- qualifiers of nouns (adjective), qualifiers of verbs (adverb) and qualifiers of qualifiers.

## **3.2 Phrase Types**

Words are grouped together into phrases which can then be joined to form sentences. A sentence constitutes different types of phrases.

### **3.2.1 Noun Phrase**

A noun phrase contains information about the noun. Normally this noun is a participant of the action described by the verb.

### **3.2.2 Verb Phrase**

These phrases generally organize words in the sentences which are syntactically dependant on the verb.

### **3.2.3 Adverbial Phrase**

An adverbial phrase is a group of related words which play the role of an adverb.

### **3.2.4 Adjectival Phrases**

An adjective phrase is a group of words in a sentence that functions in the same way a lone adjective would. Adjectives are used to modify nouns or pronouns. They give an additional detail about the meaning of a noun. In the sentence below ,underlined phrase is an example of adjectival phrase.

### **3.2.5 Postpositional Phrase**

In Malayalam language, words that serves the role of prepositions comes after the dependant noun phrase. They are called postpositions. These postpositions indicate semantic relation ship between their complements and appearing context.

## **3.3 Malayalam sentences**

A sentence is a group of words that makes complete sense. There are four types of sentences available in Malayalam based on behavior. Based on construction they are of three types.

### **3.3.1 Sentence Classification Based on Behaviour**

Malayalam sentences are basically of four types based on their behaviour.

Assertive Sentence -This is a sentence that makes a statement.

Interrogative sentence -This asks a question.

Imperative sentence -This expresses a command, a request or a prayer.

Exclamatory Sentence

To express strong feeling, extreme happiness, sorrow or wonder these kind of sentences are used.

### **3.3.2 Sentence Classification based on Construction**

Sentences in Malayalam can be of three types based on the construction- simple, complex and compound.

Simple sentences- This type contains only one main clause.

Complex sentences-These sentences contain one principal clause and any number of subordinate clauses.

Compound sentences -Compound sentences have any number of principal clauses.

Table 1. Phrase Tags

Tag	Description	Tag-Head
NP	Main Noun Phrase	Noun
VP	Verb Phrase	Verb or Auxiliary verb
NP_LOC	Noun Phrase (Karaka Phrase)	Noun with Location case
ADVP	Adverbial phrase	Adverb
ADJP	Adjecterial Phrase	Adjective
PP_INST	Prepositional phrase	Preposition

Most of the sentences in a Malayalam documents are compound or complex sentences. Hence to separate the phrase chunks first of all clauses are to be separated. Then from each clause, chunks are identified and labelled.

For each chunk type there is a head which is often the right most element of the chunk. Chunks in Malayalam are basically of five types –VP, NP, ADVP, ADJP and PP. Table1 lists few chunk types, simple explanation and tag head.

#### 4. HUMAN IMMUNE SYSTEM

Artificial Immune Systems are adaptive systems inspired by the theories of human immune systems. These principles are used for solving computational problems from Mathematics, Engineering and Information Technology [15].

Even though human body is exposed to large number of infectious agents or antigens, only few of them result in diseases. Antigens generally encompass foreign agents such as molecules on the surface of pathogens. A wide range of organisms belonging to bacteria, virus, fungi, protozoan, helminthes etc. which could cause diseases in humans are called pathogens. Pathogens can enter our body by various means, multiply and interfere with normal vital activities resulting in morphological and functional damage. Also human body reacts to harmless agents such as foreign implants, blood transfusion, pollen and drugs. The attack from the external invading agents is defended by the immune system present in our body itself. The human immune system constitutes two layers of defense viz. the innate immune system and Adaptive immune system. The nature of immune system is regulated by the activity of White Blood Corpuscles (WBC) or the leucocytes [20].

Innate immunity is an antigen –nonspecific defense mechanism that a host uses immediately or within several hours after exposure to almost any microbe. This immunity system does not recognize every possible antigen. Instead it is designed to recognize molecules shared by groups of related microbes that are essential for the survival of those organisms. The innate immune responses do not improve with repeated exposure to a given infection.

If the pathogens manage to cross the initial barriers, chemical barriers like low pH, hydrolytic enzymes and antibacterial defense mechanisms are activated. Micro flora that covers the body forms another line of defense. When the first level of barrier fails to stop the invading pathogen, it causes local infection.

Adaptive Immune System is pathogen specific. It is characterized by memory. When human body encounters a pathogen for the first time, it produces a response called primary response. If

the same pathogen stimulus repeats, a highly intensified secondary response elicits. These two responses are carried out with the help of two lymphocytes present in our blood namely B-lymphocytes and T-lymphocytes.

T-Cells or T-lymphocytes belong to the group of white blood cells known as lymphocytes and play a central role in cell mediated immunity. They have special receptor on their cell surface called T-cell receptors. T-cells are generated in the thymus but not all of them mature. Positive selection selects T cells that are capable of recognizing self antigens and Negative selection selects T cells that bind too strongly to non self antigens.

B cells are lymphocytes that play a large role in the humoral immune response. The principal functions of B cells are to make antibodies against antigens, perform the role of antigen-presenting cells (APCs) and eventually develop into memory B cells after activation by antigen interaction. B-cell development commences in the fetal liver and continues in the bone marrow. The immature B-cells undergo changes inside the bone so that they can distinguish the self cells from the non self. Receptors on the surface of mature B-cells respond to specific antigens and become activated on recognition.

The cells of the immune system make antibodies to invading pathogens like viruses. Some of the activated T and B cells differentiate to form memory cells. They continue to circulate within the body and guarantee protection from the same antigen that caused its generation.

In the AIS based phrase chunker, phrase identification is done with the help of artificial immunity system. This is implemented using the T-cell generation and maturity principle of human immunity system.

## 5. ARTIFICIAL IMMUNITY SYSTEM BASED PHRASE CHUNKER

This system is developed for identifying and separating nonoverlapping phrase chunks from sentences of Malayalam documents using the principles of human immune system. Artificial Immune System, a branch of AI has number of applications such as pattern recognition, anomaly detection, robotics, and data mining [11].

Table 2.POS Tags

Tag	Description	Tag	Description
NOUN	Noun	AdvP	Adverbial participle
ACC	Accusative	Adv	Adverb
DAT	Dative	Adj	Adjective
GEN	Gentive	AdvT	Adverbial clause of Time
PSP1	Postposition Type1	Adv-S	Adverbial clause of Supposition
VERB	verb	Vaux	Auxiliary Verb

### 5.1 POS Tagger

No tagged corpus or tag set is available for Malayalam language. To prepare the tagged corpus first of all, POS tag set with 52 tags is developed. Few POS tags are shown in the table 2. Documents from Malayalam dailies, magazines and textbooks have been selected and tagged using the POS tagger. POS tagger was designed using the probabilistic approach Extended Conditional Random Field. This tagger exploited the morphological, syntactical and semantical features of Malayalam language.

Working of this tagger in fig.1 is as follows. The input document is divided into tokens. Then each token is sent to the word analyzer for the detailed analysis. First it checks the token to decide whether it is a compound word or not. If it is a simple word the local information is collected from the lexicon. Else the token is sent to the compound word splitter to find out the morphological details and the constituents of it. The tagger assigns to each token, all possible POS tags with the help of local information provided by the word analyzer. Then to resolve the ambiguity, Extended Conditional Random Field model is used with contextual information and it eliminates all but one tag.

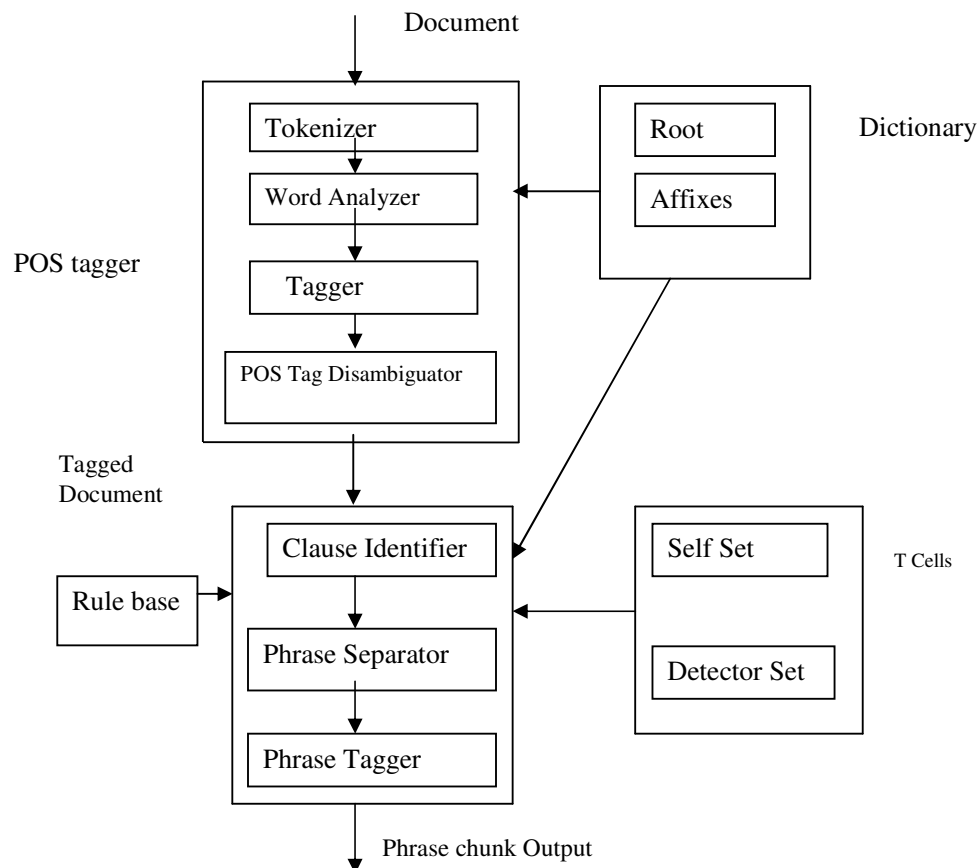


Fig.1 Detailed Architecture of Phrase Chunker

## 5.2 Clause Identifier

Malayalam sentences are mostly complex or compound sentences containing multiple clauses. To identify phrase chunks it is an essential step to identify and separate the clauses in a document. Clause identifier identifies and separates clauses from the sentences using a rule base and gives the output to the chunker.

To obtain the clauses it is necessary to find out where a clause begins and where it ends. The POS tags assigned to each and every token in the sentence is used to determine these positions. Normally main clause ends with a verb and a subordinate clause ends with an infinite verb. This module identifies clauses with the help of handcrafted linguistic rules. A clause may end either with a verb, Auxiliary verb, Adjectival participle or an Adverbial participle.

In the rules given below, N, N+1, N+2, N+3 are the positions of the current word and the three words following the current word. Examples of rules

Rule 1: IF N+1 adjp N+2 Noun then separate clause

(if (N+1)<sup>th</sup> word is adjectival participle and (N+2)<sup>th</sup> word is noun then it is a adjectival clause, hence separate the clause)

Rule 2: IF N+1 adjp N+2 adjp N+3 Noun then separate clause

Rule 3: IF N+1 advp then separate clause

## 5.3 Chunker

Each clause obtained in the previous step is taken one by one and sent to the chunker. All the phrases corresponding to each clause is identified and separated. Then they are labelled with phrase tags.

### 5.3.1 Methodology

Chunker is implemented using the principles of human immune system. There are two levels of defence mechanism in human body. First level of defence called Innate Immune system is available since birth and is regulated by white blood cells. Innate immune system defends the external organisms by making barriers such as mucus, low PH, saliva etc. Second level of defence is called Adaptive Immune system. This defence is pathogen specific and is controlled by Lymphocytes constituted of T-cells and B-cells. When a specific pathogen enters the human body the T-cells with the specific receptors recognizes and tries to destroy them. In case, the number of cells to attack the invading pathogen is insufficient then they are multiplied by clonal selection and attack the pathogen [9].

In a phrase chunker a POS tagged document is the input. Using the phrase grammar the POS string can be analyzed to identify the phrase chunks. In Artificial Immunity based system the POS string can be assumed as a pathogen. A pathogen can be identified by the corresponding T-cells. When the chunker receives a phrase chunk with the POS tags, the POS tag string is detected by the T-cells created in the beginning. Since the number of valid phrase chunks identified for Malayalam language is twenty, there are 20 different T-cells to identify these phrase chunks. T-cells which recognize these valid phrase chunks are called self cells and others



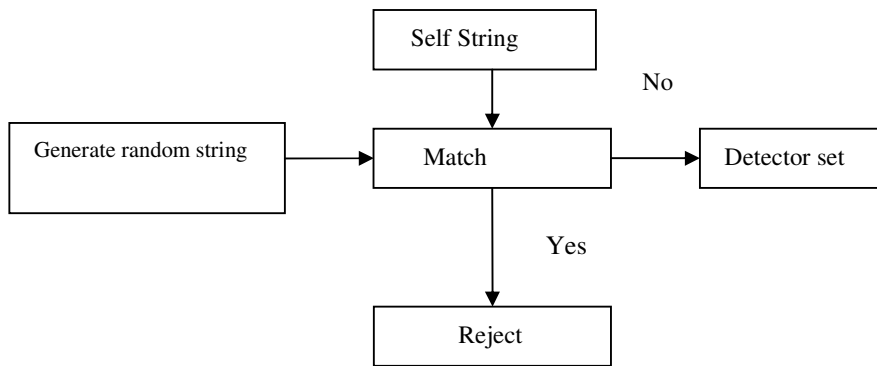


Fig2 Self/ Nonself cell preparation

are called non self cells. As shown in fig.2, the first step of this system development is to generate all valid strings of POS i.e. collection S of self strings. The second step is to generate random strings (collection R) and then match the strings of R against all strings of self, string that match self are eliminated. Strings that do not match any of the strings of S become members of D (Detector set). Self set and detectors set is created to perform the job of the T-cells. Table 3 gives some examples of generated POS patterns. Combinations of POS's that are aright according to phrase grammar are of self set, while the other combinations form the non self [20].

Table 3 Examples of Generated Patterns

Self Set Patterns	Detector Set Patterns
Adj Noun	Noun Verb
Adj Nounlist	Advp Noun
Adjp Noun	Adjp Verb
Verb Vaux	Verb Karaka
Advp verb	Vaux Noun
	Karaka Verb

Once T-cells are created, the system is ready to work like an artificial immune system. ie the T-cells are mature enough to detect the presence of incoming phrase chunks (pathogens).If the chunks are detected, they are separated (thrown out) from the incoming clause.

Chunker module accepts a stream of words with the corresponding POS tags as input and group sub-sequences of tags that most likely form a phrase. Every subgroup of tags or phrase is compared to every pattern in the Detector set. If there is a match found then the tested pattern is of non self and a process is activated to reject the incoming phrase. If there is no match then comparison is carried out with the self set. If the tested pattern matches any self pattern then "accept" process is activated. This action is continued until all the clauses are split into phrases. Working of chunker is described in fig.3.

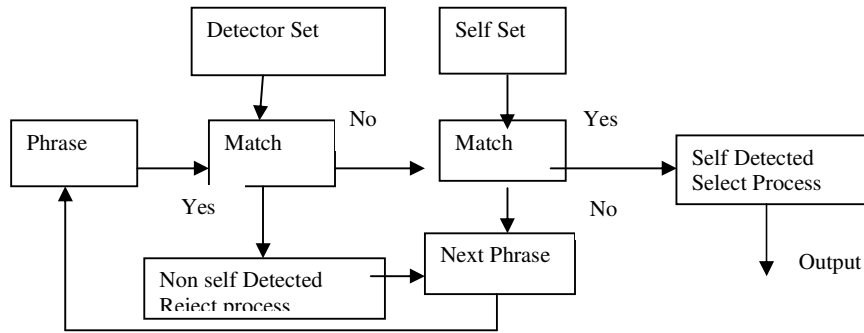


Fig3 Working of Phrase Chunker

## 6. RESULTS AND DISCUSSION

Phrase chunker is designed and implemented using J2SDK1.4.2 and MySQL. Its performance is evaluated using standardized techniques precision, recall and F-score where Precision is defined as a ratio of number of correct chunks to the number of chunks in the output and recall is the Ratio of number of correct chunks to the number of chunks in the test data.

Table 4. Performance of the AIS based Phrase chunker

Chunk	Precision	Recall	F-Score
NP	93.5%	92.6%	93.0%
VP	96.7%	97.1%	96.9%
NP-Obj	92.0%	92.2%	92.1%
NP-Dat	96.4%	97.0%	96.7%
AdvT	94.3%	95.2%	94.7%
AdvR	91.0%	90.5%	90.7%
AdjQ	87.9%	88.3%	88.1%
PP-Time	95.0%	95.0%	95.0%
PP-Inst	92.8%	91.4%	92.1%

F-score =  $2 * \text{recall} * \text{precision} / (\text{recall} + \text{precision})$  [21]. Precision, recall and F-score obtained for various types of chunks are shown in table4.

Documents related to five different fields are selected as test corpus. All types of sentences are tested. Average precision is 96% and recall 93%. Out of 5023 phrase chunks tested, 4671 chunks gave the results correctly.194 failure cases are identified where phrases were not in

proper order and also overlapping phrases were present. To separate the phrases correctly, the word order within the phrase has to be corrected.

✓ Till now no phrase tagger or tag set was available for Malayalam language. This is the first attempt towards a shallow parser for this language.

✓ Nobody has tried artificial immune System for the development of a phrase chunker.

✓ We have exploited the self, nonself cell properties and memory cell working in our system. This can be further improved to incorporate the cloning principles and affinity maturation.

This phrase chunker is less complex and has high speed due to the reduced number of computations. Without much difficulty, system can be updated just by varying the self/non self cell patterns.

## 7. CONCLUSION

Natural language understanding is an important field of AI, concerned with the speech and language understanding between human and computer. Understanding human language is a difficult task due to the ambiguity of its structure. Also exact parsing of natural languages is extremely complex. QA systems, another application of NLP requires retrieval of exact answers lying somewhere in the document. Accessing the correct information is again language dependant. Chunking can partially resolve these ambiguities. Phrase chunking can be used as a preprocessing tool in all NLP applications. Chunking produces a certain level of grammatical information useful for named entity recognition which intern provides assistance in information extraction, retrieval and QA systems.

We have presented a through investigation of using Artificial Immune Systems for phrase chunking in Malayalam. Phrase tag set with 20 tags is also developed. We could present chunking problem as a defense towards antigens, hence it become possible to apply immune system theory for phrase chunking.

## REFERENCE

- [1] Stefan Schwarzler Joachim Schenk, Frank Wallhoff, Gunther Ruske(2008). Natural Language Understanding by Combining Statistical methods and Extended Control Free Grammars. *Proceedings of 30<sup>th</sup> DAGM Symposium on Pattern Recognition, Springer-Verlag Berlin, Heidelberg.*
- [2] Nguyen Tuan Dang ,Do Thi Thanh Tuyen(2009). Natural Language Question Answering model Applied to Document Retrieval System. *World Academy of Sciences, Engineering and Technology 51, 2009.*
- [3] Yoshimasa Tsurunoka ,Junichi Tsujii, Sophia Ananiadou .Fast full Parsing by Linear Chain Conditional Random Field(2005). *Proceedings of 12<sup>th</sup> conference of the European chapter of the ACL pp 790-798.*
- [4] S. Abney: Parsing by chunks. In R. Berwick, S. Abney, and C. Tenny. Principle based Parsing. *Kluwer Academic Publishers (1991).*
- [5] L. A. Ramshaw and M. P. Marcus(1995). Text chunking using transformation-based learning. *Proceedings of the Third ACL Workshop on Very Large Corpora (1995).*
- [6] Soheila Kiani, Tara Akhavan and Mehrnosh Shamsfard(2009). Developing a Persian Chunker Using a Hybrid Approach. *Proceedings of International Conference on Computer Science and Information Technology, pp 227-233.*
- [7] A .R .Rajarajavarma. Keralapanineeyam. *National Book Stall, Kottayam, 2000.*
- [8] Prof. K.S.Narayana Pillai(1995) Adhunika Malayala Vyakaranam. *Kerala Bhasha Institute, Thiruvananthapuram*
- [9] Stephani Forest, Alan S Perelson, Lawrence Allen, Rajesh Cherukuri(1994). Self-Nonself Discrimination in a computer. *In Proceedings of 1994 IEEE Symposium on Research in security and privacy.*
- [10] Dong Hwa Kim, Kye Young Lee(2002) Neural Networks Control by Immune Network Algorithm based Auto weight Function Tuning. *0-7803-7278-6/02/\$10.00 @ 2002 IEEE.*

- [11] Ching Ham Tejhanta Singh ,Sivashankar.B Nair(2005)An Artificial Immune system for a Multi agent Robotics System.*World Academy of sciences Engineering and Technology 2005*.
- [12] Nasser Omer Sahel Ba-Karait,Site Mariyam Shamsuddin and Rubita Sudirman (2009). Swarm Negative Selection Algorithm for Electroencephalogram Signals Classification.*Journal of Computer Science 5(12) pp 998-1005,2009*.
- [13] Leandro N De Castro,Fernando J Van Zuben(2002).Learning and Optimization using the Clonal Selection Principle. *IEEE Transactions on Evolutionary Computation Special Issue on Artificial Immune System vol6 No.3 pp 239-251,2002*.
- [14] Fangjia Li,Shangee Gao,Wee Wang and Zheng Tag(2007).An Adaptive clonal Selection selection Algorithm for Edge Linking Problem.*IJCSNS International Journal of Computer Science and Network Security vol 9 , No.7 July 2007*.
- [15] Akshat Kumar and Shivashankar B Nair(2007).Artificial Immune System Based Approach for English Grammar Checking.*LNCS pp348-387*.
- [16] Yong -Han Lee ,Mi- Young Kim and Jong-ynok Lee(2005). Chunking using CRF in Korean Texts. *IJCNLP 2005,LNAI 3651 PP-155-164,2005*.
- [17] Hiroshi Echizen-ya,Kenji Araki(2010). Automatic Evaluation Method for Machine Translation using Noun Phrase Chunking.*Proceedings of 48<sup>th</sup> Annual meeting of the ACL, pp- 108-117,Uppsala, Sweden,july 2010*.
- [18] Ruy L Milidui Cicero Nogueira Dos Santos ,Julio C Darle(2008).Phrase Chunking using Entropy Guided Transformation Learning.*Proceedings of ACL-08 HLT, PP-647-655*.
- [19] Dr.C.K.Chandrasekharan Nair.Adisthana Vyakaranam. *Kerala Bhasha Institute, Thiruvananthapuram, 1997*
- [20] Simon M Garrett(2005).How do we evaluate Artificial Immune System. *Evolutionary Computation 13(2):145-178*.
- [21] Daniel Jurafsky and James H Martin(2002).Speech and Language Processing.*AI Pearson Education Series in AI.,First Indian Print 2002*.

## AUTHORS

**Bindu.M.S** received her B.Tech degree from M.A College of Engineering, Kothamangalam in 1986 and M.E degree from Coimbatore Institute of Technology, Coimbatore in 1988.She is currently pursuing the Ph. D. degree in the research area of Natural Language Processing from Cochin University of Science and Technology,

Cochin, India. During 1988-1998 she was with Manipal Institute of Technology, Manipal, as Lecturer and then as Reader in the Department of Computer Science and Engineering. Currently she is working as Reader in the Department of Computer Applications with Mahatma Gandhi University, Kottayam India. She has published several papers in International and National conference proceedings. Her research interests include Natural Language Processing, Artificial Intelligence and Information Retrieval.

**Dr. Sumam Mary Idicula** took B.Sc (Engg) degree in Electrical Engineering from College of Engineering Trivandrum in 1983. She pursued her Master studies in the field of Computer and Information Science in Cochin University of Science & Technology and took M.Tech degree in 1986. She started her carrier as lecturer in the Department of Computer Science of Cochin University of Science & Technology in 1987. She took PhD degree in Computer Science later and is now working as Reader in the same department.

She is an active researcher in the field of Natural Language Processing and Human Computer Interaction. She has undertaken 3 major projects supported by ISRO and UGC in the field of Natural Language Processing and 2 major projects supported by AICTE and KSCSTE in the field of Human Computer Interaction. She is guiding several M.Tech students & Ph.D Scholars. About 40 research papers have been published by her in the field of Computer Science in reputed journals and in international conferences. She has visited Europe and United States for participating in InternationalConferences&Workshops.

She is a member of the Board of Studies of Computer Science and Board of Studies of Computer Applications of Cochin University of Science & Technology and also a member of the Academic Committee of CUSAT.