IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

Grammatical Error Detection Model for Assamese Sentences

Hirakjyoti Sarma[1], Debasish Das[2], Kishore Kashyap[3]

Student, Department of Information Technology, Gauhati University, Guwahati, India [1, 2]

Assistant Professor, Department of Information Technology, Gauhati University, Guwahati, India [3]

Abstract: Assamese is an Eastern Indo-Aryan language used mainly in the state of Assam. This paper presents an introduction to the Grammatical analysis model of the structures of Assamese sentences which has evolved from extensive computational, linguistic, and psycholinguistic research, provides a simple set of rules for describing the common properties of all natural languages and the particular properties of individual languages. .Parsing is important in Linguistics and Natural Language Processing to understand the syntax and semantics of a natural language grammar. Parsing natural language text is challenging because of the problems like free word order, ambiguity and inefficiency. We proposed one model which is based on Top down parsing method and a context free grammar (CFG) for Assamese language with some limited domain.

Keywords: Assamese, Grammar Checker, CFG, Parsing.

I. INTRODUCTION

Natural language can be describe by formal language, in 1950 Noam Chomsky first attempted to give a precise characterization of the structure of natural languages, he tries to describe the syntax of a natural language in terms of mathematical rules. Which is later known as Context free grammar (CFG).Where grammar rule are represented by production rules and those grammar or production rules allow us to write a parser to check whether an input string is grammatically correct or not.

We have consider some simple Assamese sentence and design production rule for them .Using left factoring we remove left recursion problem of our proposed production rules. We have taken the top down parsing scheme non recursive predictive parsing algorithm to parse our proposed grammar for Assamese language. We have taken tagged sentences as input of our designed parser.

II. A BRIEF OVERVIEW OF ASSAMESE LANGUAGE

The Assamese language is cordially associated with the most important Indo-European Language Group. We have to study the Indo-European Language group to find the origin of Assamese language though it seems that it is made up of the Prota-Astrolied and Sino-Tibetan Language Group. Ascoli divided the Indo-European Language group into two main group viz. Satam and Centum. Indo-Aryan Languages are derived from the Indi-Iranian group which is one of the four sub-division of Satam. Assamese Language has also come through the three stages [(1) Old Indo-Aryan 1500BC-600BC→ (2) Middle Indo-Aryan 600BC-1000AD→ (3) New Indo-Aryan 1000AD-till now] of Indo-Aryan Language as the other Modern Indian Languages. The Indo-Aryan Languages, viz. Assamese, Bangla, Oriya etc., are derived from Avahattha through MagadhiApravhransa. The earliest evidence of Assamese dates back to the literature of the Charyapadas, written by a few Buddhist scholars.

The Assamese language present in the Charyapadas, reflects the initial stages of the evaluation of the Assamese language. There are eight vowel phonemes and twenty-one consonant phonemes including two semi-vowels in Standard Colloquial Assamese.

III.METHODOLOGY

DESIGNING CONTEXT FREE GRAMMAR FOR ASSAMESE LANGUAGE:

G = (∑, V, S, P) Where ∑ is a finite nonempty set called the terminals, which will be parts of speech for Assamese language. For our proposed parser we consider six types of parts of speech. They are Noun, Pronoun, Article, Adjective, Conjunction, and Verb.

S →NPVP|NP
NP → noun conjunction noun|nounART|Noun pronoun conjunction pronoun|noun pronoun noun|nounpronoun|noun|noun noun conjunction pronoun|nounnoun|noun pronoun pronoun|nounadjective|noun pronoun adjective|noun noun conjunction noun|noun noun adjective| pronoun conjunction noun||Pronoun ART|pronoun pronoun conjunction pronoun|pronoun pronoun noun|pronounpronoun|pronoun|pronoun noun conjunction pronoun|pronounnoun|pronoun pronoun pronoun|pronounadjective|pronoun pronoun adjective|pronoun noun conjunction noun|pronoun noun adjectiveVP → noun verb|noun noun verb verbnoun|noun verb verbnoun|noun verb verb|noun verb noun|pronounverb|pronoun noun verb verbnoun|prooun verb verbnoun|pronoun verb verb|pronoun verb noun|verb|verbadjective|verbnoun|adjective noun verb pronoun|adjective noun verb

Fig 1: Proposed production rules for context free grammar for Assamese language

DOI 10.17148/IJARCCE.2015.411118

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

S→NP S1 S1→VP/ ϵ NP→NP5 NP1/adjectiveNP3 NP1→conjunction noun/ART/ ϵ /AP/noun NP4/pronoun NP2 NP2→conjunction pronoun/NP5/ ϵ /adjective NP3→ ϵ /adjective/NP5/conjunction adjective NP4→conjunction NP5/ ϵ NP5→noun/pronoun	AP→adjectiveAP1 AP1→ ϵ /conjunction adjective VP→NP5 VP1/verbVP3/adjective noun verb VP4 VP1→verbVP2/noun verb verb noun VP2→verb VP5/VP5 VP3→ adjective/VP5 VP4→pronoun/ ϵ VP5→noun/ ϵ

Fig 2: Production Rule after Left Factoring

First(S)= {noun,pronoun,adjective} First(S1)= { verb,adjective,noun,pronoun , ϵ} First(NP1)={conjunction ,ART, ϵ,adjective,noun,pronoun} First(NP2)={conjunction , ϵ,adjective,noun,pronoun} First(NP3)={conjunction , ϵ,adjective,noun,pronoun} First(NP4)={conjunction, ϵ } First(NP5)={noun,pronoun} First(NP)={noun,pronoun,adjective}	First(VP)={verb,adjective,noun} First(VP1)={verb,noun} First(VP2)={verb, ϵ,noun} First(VP3)={adjective,noun, ϵ } First(VP4)={pronoun, ϵ } First(VP5)={noun, ϵ } First(AP)={adjective} First(AP1)={ ϵ,conjunction}
Follow(S1)= Follow(S)={$} Follow(NP)= Follow(NP1))=Follow(NP2))= Follow(NP3)=Follow(NP4) = Follow(NP5)= { verb,adjective,noun,pronoun,conjunction ,ART ,$}	Follow(AP)=Follow(AP1) Follow(VP)= Follow(VP1)= Follow(VP2)= Follow(VP3)= Follow(VP4)= Follow(VP5)={$}

Fig 3: first and follow of our proposed grammar

Stack	Input	Production rule
$S	Pronoun noun verb$	
$S1 NP	Pronoun noun verb$	S-> NPS1
$S1 NP1 NP5	Pronoun noun verb$	NP->NP5NP1
$S1 NP1 Pronoun	Pronoun noun verb$	NP5->Pronoun
$S1 NP1	Noun verb$	
$S1 NP4 noun	Noun verb $	NP1->noun NP4
$S1 NP4	Verb $	
$S1	Verb $	NP4-> ϵ
$VP	Verb $	S1->VP
$VP3 verb	Verb $	VP->verb VP3
$VP3	$	
$	$	VP3-> ϵ

Fig 4: Parsing of "মইভাতখালোঁ"

	Noun	Pronoun	Verb	Adjective	Art	Conjunction	$
S	NP S1	NP S1		NP S1			
S1	VP	VP	VP	VP			ϵ
AP	adjectiveAP1						
AP1	ϵ	ϵ	ϵ	ϵ		conjunction adjective	ϵ
NP	NP5 NP1	NP5 NP1		adjectiveNP3			ϵ
NP1	noun NP4	pronoun NP2	ϵ	AP	Art	conjunction noun	ϵ
NP2	NP5	NP5		Adjective		conjunction pronoun	ϵ
NP3	NP5	NP5	ϵ	Adjective		conjunction adjective	ϵ
NP4	ϵ	ϵ	ϵ	ϵ		conjunction NP5	ϵ
NP5	Noun	pronoun	ϵ				
VP	NP5 VP1	NP5 VP1	verbVP3	adjective noun verbVP4			
VP1	noun verb verb noun		verbVP2				ϵ
VP2	VP5		VerbVP5				ϵ
VP3	VP5		Verb	Adjective			ϵ
VP4		pronoun					ϵ
VP5	Noun						ϵ

Fig 5: predictive Parsing Table for our proposed Assamese grammar

DOI 10.17148/IJARCCE.2015.411118

For developing production rule for Assamese language we consider 31 types of subject and 16 types of predicate.

Using left factoring we eliminate left recursion of our developed grammar. After eliminating left recursion we rewrite our production rules as shown in Fig 2.After that we computing FIRST and FOLLOW for our proposed grammar using standard rules for computing first and follow ,which results as shown in Fig 3.With the help of FIRST and FOLLOW we Create Parsing table using standard rules as shown in Fig 5.

Non recursive Predictive parsing algorithm

Let a be the first symbol of w and X be the top stack symbol;

While (X!= $){ if(X = a) pop the stack and let a be the next symbol of w;

else if (X is a terminal) error();

elseif (M[X,a] is an error entry) error();

elseif (M[X,a] = X→Y1Y2 ……Yk){ output the production X→Y1Y2 ……Yk ;

pop the stack;

push YkYk-1, …….,Y1 onto the stack, with Y1 onto the stack ,with Y1 on top;}

Let X be the top stack symbol ; }

if X=$,Sentence is Accepted.

IV. EXPERIMENTAL RESULT

We test our parser for subject part , between 31 inputs our parser will able to parse 27 types of input sentence which is 87.09% of total .So based on subject part we can say that implemented parser is 87.09% efficient.

Successful parsing of an Assamese sentence"মইভাতখালোঁ" which is in the form of "Pronoun-Noun-Verb" is shown in Fig 3

V. CONCLUSION

This paper presents a top down predictive parsing approach to parse simple assamese sentences .we only consider only six types of parts of speech while developing our proposed grammar for assamese sentences . In future work we have to consider as many structures of Assamese sentences as we can and some more parts of speech.

REFERENCES

[1] Deka and kalita.," Adhunik rosona bisitra," 9th edition, assam book dipo , 2007

[2] Satyanath Bora.,"Bohol biyakoron," 1st edition,bina library, 2012

[3] Utpal Sharma" Unsupervised Learning of Morphology of a Highly Inflectional Language," Ph.D. thesis, Department of Computer Science and Information Technology, Tezpur University

[4] Navanath Saharia,Utpal Sharma,Jugal Kalita," A First Step Towards Parsing of Assamese Text,", 10 Dec.

[5] K. M. Azharul Hasan, Al-Mahmud, Amit Mondal, Amit Saha," RECOGNIZING BANGLA GRAMMAR USING PREDICTIVE PARSER,",10 Dec 2013

[6] Monica S. Lam, Alfred V. Aho, Ravi Sethi, Jeffrey D. Ulman." Compilers : Principles, Techniques and Tools,"2nd edition, Pearson (2008)

[7] Tao Jiang, Ming Li, Bala Ravikumar, Kenneth W. Regan," Formal Grammars and Languages,",10 dec.2013

[8] Kishore Kashyap, Hirakjyoti Sarma , Shikhar Kumar Sarma," Luitspell: Development of an Assamese Language Spell Checker for Open Office Writer", vol. 2 Issue. 5, EJAET, 2015.

[9] Mirzanur Rahman, Sufal Das and Utpal Sharma," Parsing of part-of-speech tagged Assamese Texts,",10 Dec.2013.